

7 May 1981

Reagan's need of a proliferation policy

Another bout of anxiety about the proliferation of nuclear weapons is plainly in prospect. The announcement last week by the government of South Africa that it had started to manufacture enriched uranium is certain to rekindle anxiety about the republic's long-term intentions. Dribbles of news, not easily verified, from Pakistan, seem calculated to have the same effect. And the disagreement between the United States and India about the supply of American enriched uranium for Indian reactors seems finally to have led to the proposal that the treaty between the two countries on the supply of nuclear materials, first signed in 1963, should now be abandoned. The consequence will be that the Indian Atomic Energy Commission will be able to use the reactors covered by the agreement as it wishes, and that there will be no external restraints (except the shortage of resources) on its enrichment of uranium. Taken separately, these developments are not as worrying as they may seem to some. A more substantial cause for anxiety is that they coincide with the Reagan Administration's failure to work out its policy on nuclear proliferation.

One difficulty, not as far beyond President Reagan's control as it may seem in what has been called his second honeymoon, is that the nomination of Mr James Malone as the assistant secretary at the State Department responsible for the non-proliferation policy is held up in the Senate. That, however, need not mean that nothing can be decided. The most urgent need is that the new Administration, which cannot welcome being bound (as it still is) by the terms of President Carter's Anti-Proliferation Act, should decide whether it wishes to amend the act, and how, or whether it should simply repeal it. The essential folly of the act is that it seeks to apply to all potential purchasers of nuclear materials from the United States a unilateral policy which overrides pre-existing agreements (as in the case of India), even with those states that have ratified the non-proliferation treaty (which India has not). The act is also closely linked with President Carter's interdiction against the reprocessing of nuclear fuel for the extraction of plutonium. Applying this rule to all buyers of United States uranium was scarcely palatable when the United States itself was not reprocessing spent uranium. This policy will be less defensible if, as seems inevitable, reprocessing of fuel begins again in the United States in the next few years. But even deciding what should be done domestically is hampered by the Administration's ill-preparedness. Thus the leaderless Nuclear Regulatory Commission is at sixes and sevens (see *Nature* 30 April, p.727), and unlikely to be able to contribute much to constructive policy formation.

Another imponderable is that nobody can tell how Congress would respond to suggestions that present policies should be modified. President Carter's men, when persuading Congress to enact the present legislation, did much (and more than they should have done) to persuade anybody who would listen that the risk of the emergence of nuclear weapons elsewhere must be simply proportional to the quantity of plutonium in existence. That message, however over-simple, may have stuck. Certainly Congress is unlikely willingly to give up the powers that act has almost fortuitously presented it with. Having wished the Non-Proliferation Act on Congress and himself, President Carter had to suffer the humiliation of asking Congress to make an exception of the provision of enriched uranium required of the United States by the 1963 bilateral agreement between itself and India. President Reagan cannot risk asking Congress to embark on a piecemeal amendment of existing legislation. He needs both a

coherent domestic policy on the development of nuclear power, with our without reprocessing, and an external policy that makes some concessions to the principle that countries overseas need to know in advance what is required of them when they buy American nuclear materials. And, even then, Congress will have to be re-educated. That, presumably, is one of Mr Malone's intended tasks.

So what should the Reagan Administration require of its nuclear customers overseas? The Non-Proliferation Act requires that potential customers should sign the Non-Proliferation Treaty (which previously unwilling Egypt has now done). In that respect, the United States is asking more of its customers than other suppliers of nuclear materials, France for example. There is, however, nothing wrong with that provision — would that other suppliers made the same demands. The iniquity of the act is that even those customers who have ratified the treaty, and thus bound themselves both not to carry out nuclear explosions and to open all nuclear installations to the international inspectors, are required not to extract plutonium from their spent fuel. (Technically, the act allows reprocessing if the objective is to obtain plutonium for research on fast reactors, but only then with the prior agreement of the United States.) President Carter appeared not to have appreciated how offensive were these provisions overseas. Some states can fairly argue that they need to reprocess spent uranium for economic reasons. All non-nuclear powers among them can rightly say that if they have, at the urging of the United States, ratified the Non-Proliferation Treaty, it is inequitable that further restrictions should then be arbitrarily imposed. Can it be, they say, that the United States does not believe in the international system of safeguards that it has helped to create? Worse still, does the United States assume that states bound by international treaties will nevertheless break their word whenever they have the chance?

The plain truth, of course, is that there is no necessary connection between the amount of plutonium within a country's borders and the chance that it will go on to make nuclear weapons. There is plenty of evidence that the factors that drive states to one policy or another are political, not technical. Except for the five well-advertised nuclear powers, political considerations appear so far to have argued for restraint and propriety. Israel has gone further along the nuclear road than most other non-nuclear states, and has been compelled to do so in a significant way. The word has been put about that nuclear technology in Israel is advanced, and that there may be plutonium sufficient for several weapons tucked secretly away. So Israel is a kind of crypto-nuclear power. The snag (from the Israeli point of view) is that it would be exceedingly dangerous to carry out a nuclear test, or openly to confess that there is a stock of bombs. For then the wrath not merely of the Middle East but of the super-powers would be visited on the Israeli government. It is not unreasonable to guess that South Africa may be dicker with a similar policy. The Indian nuclear explosion in 1974 (which prompted the still unanswered question of when the second explosion will take place) had more ambiguous motives, but served quickly to remind the Indian government of its dependence on economic assistance from overseas.

The Reagan Administration is as well placed as its predecessors to manipulate political pressures such as these, and thus to inhibit the proliferation of nuclear weapons. Domestic politics would probably prevent the Administration from demanding that Israel should sign the Non-Proliferation Treaty (now that Egypt has



done so) if it is to continue receiving assistance with conventional armaments. Pakistan would be an easier target, however — and Pakistan's signature of the treaty would go a long way to persuade India that playing about with bombs is not worthwhile. The strangest feature of President Carter's advocacy of the Non-Proliferation Act of 1978 is that, although used to decrying what are called "technological fixes" in other contexts, he relied on just such a device in the battle against proliferation. The question

now is whether President Reagan will be more subtle. He will have to persuade Congress that the American influence on the pace of proliferation is best left to him and to the State Department, not left to the wooden interpretation of an over-restrictive piece of legislation. At the same time, he will have to persuade both Congress and its electors that nuclear power is not necessarily the end of the world. The more time that slips by, the harder will be the task.

Advice and dissent on civil science

The inquiry by the House of Lords Select Committee on Science and Technology seems already to have unearthed a substantial problem, the constitution of the committee known as the Advisory Board on the Research Councils (see page 3). Much of the evidence so far sent in to the committee, that from the Royal Society for example, has argued that the board's interpretation of its function has been unhelpfully narrow. Constitutionally the board is advisory to the Secretary of State for Education and Science, who is in turn responsible for writing the cheques that keep the research councils alive. Since it was set up a decade ago as part of the Rothschild reorganization of civil science, the board has been almost exclusively a forum for horse-trading between the research councils. Which budgets should be increased more quickly than inflation, which should be trimmed? Given the inevitable dominance of the board by the powerful figures of the chief executives of the research councils, and the prospect that research budgets will indefinitely stagnate, the board's concentration on mere money and its distribution is understandable. The complaint is that, in the circumstances, there is no mechanism for tackling more general questions central to the administration of civil science.

The charge is valid even though the board seems recently (to its credit) to have taken on some general questions of common concern to the research councils as a group. Thus the board has begun an investigation of the relationship in the funding of research of the University Grants Committee and the research councils — and has made very little progress. There is an inquiry into graduate education and the basis on which research councils support graduate students in universities and other institutions, and also into the employment (and fate) of postdoctoral researchers not members of academic staffs. In due course, if the busy people who have committed themselves to work on these questions can find the time, the research councils may be helped to a better understanding of some important questions. But they are largely domestic questions. They do not apparently include the long overdue appraisal of the Rothschild reorganization which among other things gave operational departments charge of part of the budget for civil science. Instead, the board has lent its name to the piecemeal dismantling of the Rothschild system typified last year by the transfer back to the Medical Research Council of funds earlier transferred to the Department of Health.

The vein of discontent with the advisory board that the House of Lords has tapped is in no sense a new discovery. Ironically, however, the narrowness of the board's terms of reference is itself a consequence of the Rothschild reorganization. In 1971, the Council for Scientific Policy (the board's predecessor in the 1960s) and Lord Rothschild independently embarked on a consideration of how the British government's support for civil science should be organized. The council concluded that there should be an advisory committee (called a board) with wide powers to consider the research policies of government departments and research councils. Lord Rothschild eventually came to the opposite conclusion — that the Council for Scientific Policy should become strictly advisory, and that its chief task should be to recommend the division of the funds available between the research councils. Sir Frederick Dainton, the council's chairman at the time and already incensed that his own recommendations were kept secret while the Rothschild inquiry was under way, was understandably infuriated that the

government of the day should have accepted Rothschild's view that the committee's terms of reference should be narrow but should then have called that watered-down version of the standing advisory committee a "board", almost the only proposal of Dainton's to be accepted. The implication is that the advisory board, however unsatisfactory, has been doing this past decade what was expected of it.

The temptation now, in the House of Lords and elsewhere, will be to tinker with terms of reference. Sir Alec Merrison, the chairman for the time being of the board (and of a host of other committees) said last week that his board should have a wider remit but should not be encouraged to poach on the ground occupied by another advisory committee, the Advisory Council on Applied Research and Development. The modesty is seemingly but inappropriate. There is no reason why broader terms of reference should of themselves allow the committee to function more effectively. If it cannot get through the work it has already set itself, what reason is there to expect the same committee with a similar way of working to cover more ground effectively? At the same time, now that the board has come into being, and seems anxious (given time and resources) to tackle a variety of problems common to the research councils, it would be a misfortune if yet another reorganization should bring this good work to a stop.

Where, then, should the House of Lords look for a solution? The most glaring weakness of the advisory board is its lack of independence from the research councils as such (which is not to disparage the work of the independent members of the board). The advisory council, theoretically more influential because it reports to the Cabinet Office rather than to a department, and whose members are all independent persons, suffers from being if anything too detached from the problems that concern government from day to day. The result is, so far as the public knows, that the council has spent its time drawing attention to fields of science or technology in which research might be expected to be worthwhile. Structural questions — the relationship between the research councils and the University Grants Committee for example — seem not to have arisen. Yet these are precisely the questions on which the British government needs not merely advice but policies. The obvious solution is to leave the Advisory Board for the Research Councils much as it is, with responsibility for advice on the coordination of largely academic science, but to strengthen the advisory council. For now that the research councils have begun talking to each other in the presence of outsiders about the coordination of their work, it would be a pity to throw away the opportunities thus provided. Strengthening the advisory council is more than merely a matter of redrafting terms of reference. With the best will in the world (and even with a chairman as able as Sir Alfred Spinks), busy people meeting occasionally cannot be expected to do the hard work that circumstances require. At the least, the chairman should be a full-time appointment (in which case he or she might usefully double as chairman of the advisory board). But the committee also needs a professional staff, not mere minute-takers represented by promising young recruits to the civil service. Both changes would be easier if, as in the 1950s, responsibility for civil science were removed from the Department of Education and Science. They, and the others in prospect, would have more meaning if some British government would believe its own declaration about the importance of this subject.

Recombinant DNA guidelines wear thin

NIH moves to voluntary regulation

Washington

The US National Institutes of Health (NIH) have taken the first steps towards removing mandatory controls on the conduct of recombinant DNA research. Meeting in Bethesda, Maryland, the institutes' Recombinant DNA Advisory Committee (RAC) set up a subcommittee to discuss whether the present guidelines, which all federally-sponsored scientists have to observe, should be turned into a voluntary code of practice, and the answer is likely to be yes.

The committee's decision was in response to a proposal from two of its members, Dr Allan Campbell of Stanford University and Dr David Baltimore of the Massachusetts Institute of Technology. They suggested that the guidelines should be voluntary — eliminating the formal need for review bodies such as local Institutional Biosafety Committees — and that there should be major reductions in the containment levels recommended for various types of experiments (*Nature* 26 March, p.281).

Their proposal, which was published in the *Federal Register* on 20 March and would involve eliminating virtually all the regulatory requirements contained in Part III of the current guidelines, has generated considerable support in the scientific community.

Most members of the committee seemed to agree with the spirit of the proposition. Their willingness to relax the guidelines was illustrated in an earlier vote on reducing containment requirements for experiments involving *Escherichia coli* K12 and laboratory strain *Saccharomyces cerevisiae* host vector systems. The committee adopted the most liberal option, recommending that such experiments, making up about 80 per cent of all recombinant DNA experiments, be totally exempted from the guidelines.

Several committee members pointed out, however, that there is still considerable public concern about the potential social effects of recombinant DNA research. Hence should such research be regulated, it must be seen to be done in a responsible manner as otherwise federal or state legislation might be provoked. Other members, particularly those from the public interest community, argued against reducing the containment conditions too quickly, pointing out that although a few experiments seemed to confirm that recombinant DNA experiments posed no

greater risks than other experiments with the same organisms, this hypothesis had not yet been rigorously tested. Even Dr Campbell and Dr Baltimore agreed that experiments over 10 litres should still be subject to special review by the director of NIH, since large-scale experiments might still result in unsuspected hazards.

The committee members finally agreed that, rather than recommending to NIH director Dr Donald Fredrickson the immediate transmutation of the guidelines from formal restrictions into a voluntary code of practice, a subcommittee should look into the whole question of the future of the guidelines.

The proposal made by Dr Campbell and Dr Baltimore would result in the recommendation that most experiments involving recombinant DNA be carried out in P1 containment conditions, these representing prudent laboratory practice. Potentially hazardous microorganisms would be classified and handled according to procedures recommended by the Center for Disease Control (CDC) in Atlanta, Georgia. Several scientists have expressed concern about proposed revisions to these procedures, published by CDC last year, arguing that in certain cases they would be more restrictive than necessary. However, Dr Baltimore told the committee that he believed CDC "will eventually produce a document that is a useful guide to the appropriate laboratory practice".

Dr Baltimore also stressed that RAC would still be needed to maintain surveillance over recombinant DNA research in the future, to make decisions about

Plea to keep guidelines

A report published by the Office of Technology Assessment, after a two-year study of the impact of applied genetics, says that there are several reasons for not eliminating the guidelines completely, pointing out that "sufficient scientific concern exists for the guidelines to prohibit certain experiments and to retain containment for others", and that most experiments can already be carried out at the least burdensome containment levels.

Although conceding that NIH showed flexibility in liberalizing the restrictions as evidence showed lower risk levels than originally feared, there is criticism of the way that NIH have played the role of oversight body. Several shortcomings are listed, including the virtual non-existence of monitoring for compliance to the guidelines, and the lack of systematic evaluation of other techniques, such as cell fusion, that might present risks.

The report, which lists some of the many applications of recombinant DNA techniques in a wide range of industries, says that Congress might wish to consider the need for regulating work with all hazardous microorganisms and viruses, whether or not they are products of genetic engineering.

experiments which may present particular dangers and to help in the evaluation of ethical issues, which were going to become a very important part of a recombinant DNA activities.

David Dickson

Role of advisory committee questioned

The United Kingdom's Advisory Board for the Research Councils could provide government with advice on a wide range of public policy issues with a scientific content if its terms of reference were extended, according to its chairman, Sir Alec Merrison. At present, the board is constrained by its remit to advise the Secretary of State for Education and Science chiefly on the distribution of the science vote amongst the research councils.

The board is like a "huge torpedo" that is not being launched "at anything very much" Sir Alec told the House of Lords Select Committee on Science and Technology last month as part of its investigation into the provision of scientific advice to government. The expertise of the board's members, senior scientists and administrators, could be put to better use if it took on a role similar to the Advisory Council on Scientific Policy, which provided advice on science policy through a chief scientist to the government before its demise in 1963. Any extension of the board's remit, however, would require very careful thought, said Sir Alec, to avoid

treading on the toes of the Advisory Council for Applied Research and Development which does a good job in advising on scientific aspects of industrial policy.

Informal contact with the Department of Education and Science was generally good, Sir Alec said. But it had not worked in the case of the government's White Paper on public expenditure, published on budget day last March. Although the implications of the White Paper for the support of basic science in the universities were "profound", the board only learnt of the government's plans in a brief letter a few days before the budget statement.

Sir Alec said that by thinking that it could protect science by maintaining the level of the science vote while considerably reducing the grant for student fees, the government had clearly not understood the way in which university science is funded. The implications for that system, known as the dual support system, are so grave, said Sir Alec, that work on a report that the board is preparing into its workings has had to stop.

Judy Redfearn

Carter's science team Yesterday's men

Washington

As the Reagan Administration continues to have difficulty in filling many of its top jobs — posts which remain vacant include the directorship of the Office of Science and Technology Policy (OSTP), the chairmanship of the Nuclear Regulatory Commission, and the directorship of the Office of Energy Research — science policy officials from the Carter Administration are having mixed success in finding new positions. Most have slipped back into academic posts — the traditional fall-back for Democrats, in contrast with Republicans, who usually return to the private sector after leaving office. But for some, the prospects are still uncertain.

Most successful in finding a new job has been Dr Frank Press, Mr Carter's science adviser and director of OSTP, who was recently elected to succeed Dr Philip Handler as president of the National Academy of Sciences. Dr Press takes up his new position in Washington at the beginning of July, and has meanwhile returned temporarily to his old post as professor of geology at the Massachusetts Institute of Technology.

Of Dr Press's three original associate directors at OSTP, Dr Gilbert Omenn, responsible for health sciences, has returned to the University of Washington in Seattle, where he is professor of medical genetics, having spent the last year of the Carter Administration as associate director of the Office of Management and Budget. Dr Philip Smith, associate director for natural resources, has temporarily joined the staff of the National Science Board, the overseeing body of the National Science Foundation. The third, Dr Benjamin Hubermann, remains at OSTP keeping the director's seat warm until a successor to Dr Press is appointed.

Dr Robert Frosch, administrator of the National Aeronautics and Space Administration (NASA) under Mr Carter, is already established as president of the newly formed American Association of Engineering Societies. Dr Hans Mark, previously secretary of the Air Force and before that director of NASA's Ames Research Laboratory, was at one point tipped to succeed Dr Frosch, but has now been nominated as deputy to NASA's new administrator, Dr James Beggs.

Another academic who has returned to her former base is Dr Eula Bingham, previously Assistant Secretary of Labor responsible for the Occupational Safety and Health Administration (OSHA). Dr Bingham has rejoined the University of Cincinnati, from which she had been on leave as professor of environmental health.

In contrast, Dr Bingham's counterpart at the Environmental Protection Agency, ex-administrator Mr Doug Costle, is entering the academic world for the first time. Before his appointment under Mr Carter, Mr Costle

had been an assistant director of the Congressional Budget Office, and commissioner of environmental protection in the state of Connecticut. A lawyer by profession, he was recently named as a visiting scholar at Harvard University's School of Public Health, and a lecturer at the John F. Kennedy School of Government.

Moving in a different direction is Dr Edward Frieman, brought in to succeed Dr John Deutch as head of the Office of Energy Research. Before coming to Washington, Dr Frieman was deputy director of the Plasma Physics Laboratory at Princeton. He has now joined a Washington-based consulting firm, Science Applications Incorporated, which deals in a range of energy issues.

Two appointments have remained unchanged. Dr Donald Fredrickson, originally appointed by President Ford as director of the National Institutes of Health for a six-year term, has stayed in that position. Also staying put is Dr John Slaughter, brought in during the last months of the Carter Administration as director of the National Science Foundation partly on the understanding that Mr Reagan would not send him straight back to Washington State University, where he had only recently been appointed provost.

Less fortunate has been Dr Anthony Robbins, previously director of the National Institute for Occupational Safety and Health (NIOSH). Dr Robbins became the target of a bitter attack by conservative groups for his support of labour union positions on some occupational health issues, and was summarily removed from his post by the new Health Secretary Mr Richard Schweiker, as well as being stripped of a commission in the US Health Corps.

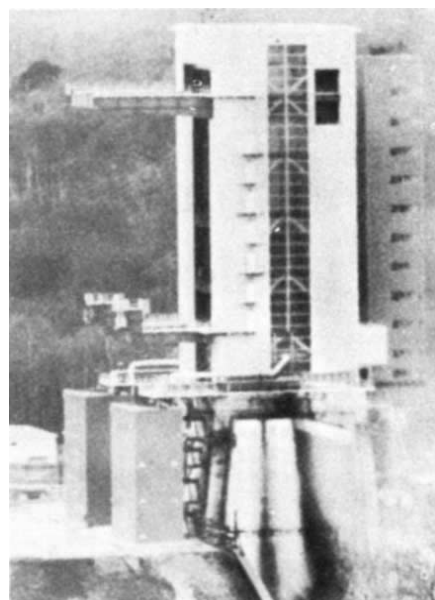
At the State Department, Thomas Pickering, previously head of the Bureau of Oceans and International Environmental and Scientific Affairs, has returned to his position as a career foreign service official. And William J. Perry, previously Under-Secretary for Defense for Research and Engineering, has joined a San Francisco investment banking firm.

David Dickson

Ariane development Launch at last?

The European Space Agency (ESA) has announced that the third test launch of the Ariane launcher, with two satellites on board, is to take place on 19 June. After extensive testing and some modifications, the problems associated with the first-stage engines have apparently been solved.

Europe's three-stage launcher programme encountered a major setback during the second Ariane test launch on 23 May 1980. Subsequent analysis revealed that instabilities in the injection system caused pressure oscillations at 2,300 Hz to develop inside the combustion chamber of one of the four first-stage Viking engines. The oscillations caused structural damage which in turn hindered cooling and resulted



Up, up, but . . . away?

in rupture of the engine. Subsequent ground tests also revealed oscillations at 2,700 Hz.

Liquid-fuelled rocket motors are notorious for such instabilities, which still confound theorists. ESA's trial and error approach to the problem has involved months of testing of alternative modifications to the injector system. One design, in which the injection nozzle diameters have been slightly widened, resulted in a complete absence of pressure oscillations during ground tests, and has now been adopted. As a further safety measure, manufacturing tolerances have been tightened.

Like its predecessors, the third Ariane is to be launched from Kourou in French Guyana. It will carry Meteosat 2, a European meteorological satellite, and APPLE, an Indian telecommunications satellite. If all goes well, the fourth and final test launch will take place in October.

Ariane's problems have hindered its progress as competitor to the United States shuttle. However, contractors such as Intelsat — the international telecommunications satellite organization — have not cancelled their options on Ariane as a result of the delays.

Philip Campbell

British aerospace Limbering up

Further evidence of the UK government's sudden eagerness to enter the space race came last week in the form of a new committee set up as a forum for the views of the space industry, users and government. This Space Consultative Committee has held its first meeting to try to determine its role.

The diverse membership of the committee, which includes representatives of government departments, the Science

Research Council, the broadcasting authorities, British Telecom and satellite and telecommunications equipment manufacturers, suggests that it is intended to coordinate views from a variety of sources, a role that in other countries might be taken by a national space agency.

Mr Keith Marshall, Under-Secretary of State in the Department of Industry, who will chair the committee, said in parliament that it would also be investigating topics such as the future programmes of the European Space Agency in telecommunications and remote sensing satellites, a satellite communications system for the Ministry of Defence and the marketing strategy of British satellite and ground-station technology abroad. He also indicated that the government would expect industry to put up all the finance for promising projects.

The committee will shortly have before it the conclusions of two reports commissioned by the government. The report of the Home Office study on direct broadcasting should be published later this month and the conclusions of the report of the Central Policy Review Staff on Britain's future in space are to be made available to the committee.

Hopes that an abbreviated version of the review body's report would be published, however, were quashed at the end of last month by a statement in the House of Commons — the government apparently fears breaching industrial confidentiality by allowing open discussion of the report. So far the only conclusions to emerge have been rather bland, such as that the United Kingdom should continue its commitment to the European Space Agency. It is to be hoped that the committee will provide a focus for wider debate. **Judy Redfearn**

European Parliament Nuclear ban debate

Brussels

A moratorium on nuclear power in Europe will be the subject of a debate in the European Parliament this week. Also on the agenda is a motion on the Geneva Appeal which was launched as a protest against France's fast breeder reactor, Super-Phénix.

The moratorium debate will focus on a report by Sir Peter Vanneck (European Democrat) who is strongly opposed to the idea. The report was adopted by the energy committee by 13 votes to 7 following a resolution by three independent Members of the European Parliament in December 1979. The moratorium would be used as an opportunity to look into the problem of storing nuclear waste, harmonizing safety standards and promoting widespread public discussion. In his report, Vanneck concedes the benefit of these points but feels that the cure would be more painful than the disease.

Depending on how widespread the moratorium would be, Vanneck calculates that by 1990 it would leave the Community short of between 45 and 115 GW, equivalent to 76 or 170 million tonnes of oil a year. By 1990, 34 per cent of the EEC's electricity needs will be supplied by nuclear power. Vanneck also points out that in 1975 electricity from nuclear power caused no deaths in the United Kingdom, while maritime transport (of oil and coal) caused 86 deaths per 100,000 employees and coal mining caused 23.4 deaths per 100,000.

The underlying cause of public concern, says Vanneck, is that plutonium produced in nuclear power stations may be used for

military needs. This issue will certainly be linked with the debate on the Geneva Appeal. In October 1978, more than 30,000 French and Swiss, including a number of leading figures, signed an appeal asking the political community in Europe at large to consider alternatives to fast breeder reactors. The appeal reflects the fears of both Swiss and French living in the neighbourhood of the 1,250 MW Super-Phénix fast breeder reactor under construction at Greys Malville in France.

The Council of Europe had held public hearings on breeder reactors in 1979. The present motion, if adopted, would call on all governments concerned to suspend work connected with breeder reactor projects. The EEC itself does not directly contribute to any fast breeder construction programme. The European Commission, however, is preparing a report on the state of the nuclear fuel reprocessing industry.

Both debates, are nevertheless rearguard actions of the anti-nuclear group in the parliament. Most parliamentarians, as previous resolutions have shown, are firmly in the nuclear camp. But the anti-nuclear minority represents nearly a fifth of the members, who are effective in keeping up pressure on the Commission and member states.

Jasper Becker

British countryside

Ever onward

The British government's Wildlife and Countryside Bill — regarded as the last opportunity for major legislation on nature conservation matters this century — is likely to generate as much controversy during its passage through the House of Commons as it did in the House of Lords earlier this year. The Lords considered a record 560 amendments to the bill and defeated the government on seven of them. Many of the arguments thrashed out then, on numerous topics relating to nature conservation, will be raised again, with the government trying to reverse some of the Lords amendments and various lobbyists trying to introduce more.

During the bill's second reading in the Commons last week, Mr Michael Heseltine, Secretary of State for the Environment, said that the government was still convinced that conservation would not be well served by compulsion. This is the chief bone of contention over clauses in the bill relating to the protection of sites of special scientific interest (SSSIs). In the Lords, the conservation lobby attempted — unsuccessfully — to amend clauses giving statutory protection to a few specially selected SSSIs to cover all such sites. The debate is likely to continue in the House of Commons committee appointed to debate the bill in detail.

The committee, which began sitting on 5 May, has 20 members, including Keith Hampson and Hector Munro for the

East Germany banks on research

High technology will be the keynote of East Germany's economic strategy for the 1980s, First Secretary Erich Honecker told last month's Congress of the Socialist Unity Party. An intensive drive to modernize and automate industry, he said, should save the country some 2,800 million man-hours during the next five years, primarily through the application of microelectronics and an increase in the number of industrial robots to be introduced into industry during the period from an original estimate of 9,000 to some 40,000.

By 1985, the country should be virtually self-sufficient in the production of microprocessors and their components. Other priority areas will include fibre optics, laser technology, industrial applications of microbiology, hydrogen and biomass energy sources and the development of new sophisticated technologies to make the maximum use of raw materials at the minimum cost in energy. (The recent shortfall and

future uncertainty of coal supplies from Poland have caused planners to pay renewed attention to the country's own lower-grade fossil fuels, mainly lignite).

Although Mr Honecker's review of the work of research establishments, from the Academy of Sciences downwards, stressed the technological and applied aspects, he did pledge the party to a systematic development of basic research "as the source of new discoveries on logical relations in nature and society". The universities, including their extension and evening courses for full-time workers, came in for special commendation as did the preliminary training courses for students at their future work-places.

The situation in the lower educational echelons, however, seems less satisfactory. Too few young people seem to be interested in scientific and technical subjects, and Honecker urged teachers and the media to make science and technology more attractive. **Vera Rich**

government and Tam Dalyell and Stephen Ross for the opposition. As well as protection for SSSIs, other controversial topics include the banning of shooting on Sundays, allowing bulls on public footpaths and how to maintain marine nature reserves.

Judy Redfearn

Reagan's cuts

Handler objects

Washington

Dr Philip Handler, in his farewell address last week as president of the National Academy of Sciences, came close to accusing the National Science Board of cowardice in its failure to protest at the Reagan Administration's cut in the budget of the National Science Foundation (NSF).

The proposal is that the foundation's budget for 1982 be trimmed to \$1,033 million, more than 30 per cent less than Mr Carter's request for \$1,353 million. At its meeting in March, the board, which is formally responsible for the activities of NSF, approved a statement on the cuts in which it complained that it had been excluded from the budget decision.

The statement also mentioned "problems of serious concern" such as the need to upgrade university research equipment and its "limited capability to reprogram funds to take advantage of emerging research opportunities". There is, however, no direct criticism of the Reagan Administration's decision as such.

Dr Handler said plaintively last week "Would that that statement were not quite so subtle". Handler accepts that the board is constrained by its formal place within the Administration, but he told the academy that the board members' "statutory responsibility to establish the priorities and allocate resources had, in fact, been usurped". And he added: "In not saying so, in unequivocal language, they may allow this incident to become a precedent".

Handler admitted that for the members of the board to have been more forthright might have meant a confrontation with the Administration in which the National Science Foundation would be the loser. "But in crafting a statement in which protest is to be found only by sophisticated reading between the lines . . . the board seemed to be party to the actions of the Office of Management and Budget."

Handler also quoted the board's contention that it was difficult to be engaged in policy decisions about budget cuts in the light of the "economic emergency", which precluded the normal discussion with the board of NSF priorities. "Emergency? What emergency?" asked Handler. "The fact that NSF was given but 24 hours, when the board was not in town, to defend its right to support social science and science education, *inter alia*, is preposterous."

David Dickson

Nuclear power in Yugoslavia Staying independent

Belgrade

Core loading is now under way at Yugoslavia's first nuclear power station in Krsko. Formally, the station — based on a Westinghouse light-water reactor — is years behind schedule, but this is in part due to an over-optimistic target of five years for construction.

The station has been built jointly by the republics of Slovenia and Croatia. The leading physics institutes of the two republics, the Rudjer Boskovic Institute (Zagreb) and the Josef Stefan Institute (Ljubljana), have been training staff for the new station while building was being completed, and working out the final safety report for the reactor.

Yugoslavia's commitment to nuclear power is strengthened by the presence of uranium deposits at Zirovski Vrh in Slovenia. There are few dissenting voices, notably hotel-keepers on the Dalmatian coast objecting to the proposed siting of the country's second reactor near the tourist resorts, and veteran anti-nuclear campaigner Dr Ivan Supek. Most Yugoslavs apparently see nuclear power as a way to greater energy self-sufficiency.



The TRIGA II reactor at the Josef Stefan Institute, used for heavy metal monitoring.

Before uranium mining and milling operations begin, however, a group from the Josef Stefan Institute is making a detailed survey of background radiation in the area. And this survey forms part of an international project to map the radioecology of the Danube basin.

Also being measured in the survey are heavy metal trace elements, including cadmium, arsenic and mercury. The area has a history of mercury mining and once the price of mercury picks up again the old mine is likely to be reopened. The mine closed in the mid-1970s and there is evidence of considerable mercury poisoning in the mine-workers involved. But studies of the effects of mining are being hampered by shortage of finance for a much-needed survey of the non-mining population in the mining area.

This highlights a general shortage of funds for research in Yugoslavia. There is no longer any central planning of science, and funds come either from direct contracts with industry, or else from a

Proton decay hunt

New Delhi

Indian and Japanese physicists have discovered three "candidate" proton decay events in a massive iron detector system set up 2,300 metres underground in the Kolar gold mines 100 km north of Bangalore in South India. Begun nearly five months ago, the experiment is to last two years during which at least six of these rare events are expected to be observed provided the proton lifetime does not exceed 10^{31} years.

The detector system consists of 140 tonnes of iron: 100 tonnes for the detector elements and the remainder for shielding. The detector elements are 4 m and 6 m long iron tubes of cross-section $10\text{ cm} \times 10\text{ cm}$. The 1,650 proportional counters are stacked to make a giant 35-layer cake that is 6 m long, 4 m wide and 4 m tall.

The on-line electronics system and the detector configuration enable the total energy of the decay process, the decay vertex and a three-dimensional plot of the tracks of decay products to be obtained. There are some 10^{32} nucleons in the iron and assuming a detector efficiency of 50 per cent, it is hoped to reach a proton lifetime limit of 5×10^{31} years.

The main background to the decay events comes from cosmic rays. Neutrinos and some muons created in cosmic ray events in the atmosphere can penetrate down to the detector; but the Kolar equipment is so deep most of the muons are absorbed in the rock above. Nevertheless, by the end of last week Kolar had recorded 223 vertical muons; 8 horizontal muons, probably coming from neutrino interactions in the rock beside the detector; three neutrino events in the detector; and three events which so far can be explained in no other way than as proton decays. However, all three events have a track which reaches (or emanates from) the edge of the chamber, so they may yet be false events, mimicked by a particle entering the chamber from outside. If the events are confirmed, the corresponding proton lifetime would be $1-3 \times 10^{30}$ years, around the number being predicted by theories which unify the weak, electromagnetic and strong interactions into one grand "gauge theory with spontaneous symmetry breaking".

K.S. Jayaraman

structure of "Self-Management Communities of Interest for Science" which administer funds voted from profits of individual commercial enterprises.

The Josef Stefan Institute gets 40 per cent of its funds from the Slovenian assembly of such communities and 60 per cent from contracts. But neither source has shown much interest in studying mercury levels in their region and monitoring a control population elsewhere. **Vera Rich**

CORRESPONDENCE

Conservation sites

SIR — Since private landownership and private incentive are among the basic tenets of Western ideology, I have to agree with Muir (*Nature*, 12 March, p.82) that the populace must suitably entice landowners if it expects them to conserve rather than utilize their land. But it is a callous and immoral science that will not use funds to intervene in natural selection or in "the very evolution of the inanimate world" if these are observed to be proceeding in a deleterious direction. Surely the spending of funds on research directed towards new forms of medical treatment, or of flood or erosion control, is not an "affront to reason", bearing condemnation, and no-one is being misled by the professed importance of these or similar interventions.

Also, thankfully, I know of no science which claims that we can study anything in nature to the limits of its information content so that "nothing new will be learned by preservation". Such would be a conceited science indeed. And, a science that attaches no value to the existence of an object, save for its information content, shows a blatant contempt for existence itself. As Santayana has written (in *Reason in Science*), "If science deserves respect, it is not for being oracular but for being useful and delightful, as seeing is".

Unique usefulness and ability to delight are also properties of jewels or "gems" and of conservation sites, although Muir, in his derisive analogy, has conveniently overlooked these. He has also ignored the one striking difference between them: the virtual indestructibility of jewels as compared with the fragility of ecological systems.

So, must we look forward dispassionately to Muir's future world with its myriad rats, biting insects and noxious weeds, with its computer data banks full to overflowing, but without museums, zoos or conservation sites? Or, should we intervene to ensure a better future where usefulness and delight remain and where Muir's biologically deficient species are not extinct before their time?

A. H. BRETAG

South Australian Institute of Technology,
Adelaide, South Australia

Literature search

SIR — PhDs normally take longer than three years to complete and their exact purpose is not agreed upon (*Nature* 22 January, p.217). As an SRC-supported research student, I would like to make a few relevant comments.

I have spent several months in my library carrying out the customary literature search. As time progresses, more and more literature has to be searched and one can only presume that the time to complete a PhD will increase! However, because I spent so long "doing" a search, I found cases of "published rediscoveries" which are attributable I think to: (1) a lack of time spent on a literature search, and (2) a language problem.

The time factor is overcome, at present, only by carrying out a minimal literature search (which in all likelihood will turn out to be inadequate) and by ignoring the language problem altogether by ignoring the literature published in a foreign language. The dilemma

facing a research student is obvious. If he/she tries to carry out a thorough literature search and learn a new language then the time spent doing this bites into those three years of support. At a recent course which I attended, which was specifically designed to teach scientists how to translate scientific Russian or German, only five people attended. The conclusion is (from albeit scanty evidence) that researchers choose to ignore the language problem.

Regarding actual literature searching it is only recently that I have found out about "computerized" abstracts, although, I'm told, they have been available for a few years. I would urge the SRC to look into the use (or lack of use) of "computerized" abstracts, if they are not doing so already, as such facilities will become increasingly important and have the potential for saving a lot of time.

I suggest that it would be beneficial if research students were given, from the outset, a course on library and information science so that they are aware of all possible sources of information. I would also suggest that it be made compulsory for research students to learn a foreign language relevant to the scientific discipline. These proposals may take up money and indeed time but to me, at least, appear a necessity if good efficient research is to be carried out. On reflection, what I have suggested favours a more "training for research" type of PhD.

ANDREW BROOKS

Department of Astronomy,
University of Edinburgh, UK

Avoiding fraud

SIR — The recent alarms about fraud in American science raise the more general problem of quality-control in research. As you correctly observe (*Nature* 9 April, p.433), peer review must be supplemented by "full and frank discussion within individual laboratories". But the process cannot stop there. When I analysed this problem in my book (*Scientific Knowledge and its Social Problems*, Oxford University Press, 1971), I remarked on the inherent difficulty of the quality-control operation in science, since there is no possibility of creating an external inspectorate for assessing the products of the complex and subtle craft skills of science.

Research is therefore necessarily a largely self-regulating activity in this respect, accomplished by peer-review and journal refereeing; but the problems of iteration, "who guards the guardians?" are then even more severe. My conclusion was that integrity and morale, at the highest professional levels, are more crucial to the health of science than perhaps in any other organized social activity.

The various moral imperatives of science, propounded by Merton, Popper and Polanyi, can thus be seen to be very relevant to the survival of worthwhile science. However, they are not so much *a priori* definitions of "science", as descriptions of attitudes and commitments whose presence and effectiveness are entirely contingent. The variability of standards of quality between fields and between milieux bears this out.

When so much of scientific information relevant to public policy is now produced not

as "public knowledge" but rather as "corporate knowhow" (in state bureaucracies or private firms), the traditional ideals and norms of science lack an appropriate social context for their reinforcement and maintenance. Under these circumstances it becomes implausible to maintain that only a prejudiced or malcontented opposition can doubt the factual veracity of any piece of technical information used as official testimony in policy debates.

J.R. RAVETZ

Department of Philosophy,
University of Leeds, UK

Cladistic clues

SIR — It is unfortunate that the attempts of Dr Miles and the Department of Public Services of the British Museum to display the most logical basic level of analysis of organisms (the level of relative organism-to-organism comparison) is misinterpreted as an attempt to coerce the general public into accepting "a fundamentally Marxist view of life". As one may perceive from the series of replies to L. B. Halstead and the cladistic literature, cladistics is a method which links many very different individual biologists working from very different sets of preconceptions. However, they seem at least to share the opinion that we learn something about organisms by studying what they are made of. The cladist may converge on the Marxist idea that events occur by "taking the form of a leap from one state to another." The cladist may similarly converge with the catastrophist creationist. However, the cladist's stand can be just the opposite of a dogmatic stand, that is, it can be an admission that there are not enough biologists to fill in all the holes in our knowledge of the history of life, and that our inadequate sample of the history of life allows us to see so little "gradualism" that one cannot dogmatically claim that it is the rule.

Halstead's point that public scientific institutions are accountable to the public is well taken. However, publicly employed scientists are accountable to the public for their salaries with ideas and interpretations which the public at large does not have time to make, as well as being accountable for actual specimen displays. Like any curator in history responsible for exhibitions, Dr Miles shares some ideas with scientists of his day, and has some ideas of his own, all of which are reflected in the exhibits he edits and presents to the public. Like any editor, he is forced to make editorial decisions; as one who studies the taxonomic group that Dr Miles has spent most of his career studying, I see no more competent individual than himself.

However, I share some of S. J. Gould's romanticism for great old halls which are closed and renovated into something that is a total stranger to me. Museums cannot help but be places of accumulation, and this catastrophistic kind of replacement of the old with the new does not seem to be the most accurate portrayal of the museum's progress in

Increased majority

In the correspondence "Majority verdict" published in *Nature* 30 April, p.730, R.L. Batten should be added to the list of signatories.

research. Perhaps the most instructive exhibition policy would be to preserve some small corner of every old exhibition hall which would allow the public to see that which their predecessors saw, and which would allow the museum to see what it stood for in the public eye. In the next generation of British Museum exhibits, maybe we will see evolution not as untestable, unscientific theory, but as a more complex though rather less testable theory than cladistic relationship, and as complex and considerably more testable than creationism.

W. KOHLBERGER

Department of Geology and Geophysics,
Yale University, New Haven,
Connecticut, USA

Cafeteria feeding

SIR — A. S. Cole (*Nature* 2 April, p.356) objects to the phrase "cafeteria feeding", applied to experimental selection of "palatable items of food" by rats. I disagree: it seems an unobjectionable, indeed mildly amusing, and appropriate piece of jargon — for once, easily understood. The only criticism might be that, in the original usage, the *palatability* of items on offer is not a *sine qua non*.

J. R. BAKER

NERC Culture Centre of
Algae and Protozoa,
Cambridge, UK

Ten years of natural history

SIR — In 1972, the trustees of the British Museum (Natural History) approved a proposal prepared by a panel of the museum's scientific staff, chaired by the then director, for a new approach to the visiting public. This proposal sought to replace the then existing exhibition which rigidly reflected the divisions of the museum into the five separate scientific departments of Botany, Entomology, Mineralogy, Palaeontology and Zoology. The new approach was timely because it coincided with, *inter alia*, changes in biology teaching in schools and an increasingly sophisticated treatment given to natural history on television.

The aims of the new exhibition scheme were summarized as follows: "To present the public with an integrated view of modern biology. All known forms of life will be included, but the new exhibition will do much more than merely demonstrate the diversity of nature. It will show how living things interact with each other and with their physical environment; it will describe the chemical and physical processes that keep them alive; it will introduce the concepts of heredity and evolution, of ecosystem and energy cycles; and it will examine man's role in the living

world¹." A long-term aim was, and still is, to include a greater number of organisms from a wider range of groups than had been shown in the past, combining both recent and fossil organisms in a single evolutionary arrangement.

The trustees have maintained a continuous and close interest in the implementation of this policy, and have conducted a careful review from which they concluded that the museum's Public Services Department has made a bold and largely successful start on implementing the long-term plan. The trustees now consider it appropriate to outline their plans for the next ten years so that those who visit and use the museum may be aware of changes and of the underlying rationale. The trustees have concluded that the scheme should proceed as shown in the table below.

The trustees believe that the implementation of this programme will be beneficial. It will, for example, allow fossil and recent mammals to be brought together in a single rational system for the first time in the museum, and it will not only give greater space to the insects but it will also relate them to the other arthropod groups.

In addition, the Department of Public

Services is developing a proposal for an exhibition devoted to British Natural History to be brought forward to the earliest feasible time. This gallery would be designed to meet the needs of committed naturalists to whom the trustees acknowledge a responsibility, although the prime commitment of the exhibitions must be to the visiting public in general.

The new stages of the exhibition proposed above will have much in common with the older galleries because they will concentrate on the museum's traditional subject matter, the diversity of living organisms. Eventually diversity exhibits are likely to comprise about two-thirds of the completed exhibition scheme. Some parts of the existing displays from, for example, British Insects and Fossil Mammals, will be used in the new exhibitions.

The trustees are also aware of the need for an exhibition related to the museum's interests in geology and have accepted in principle proposals by a Joint British Museum (Natural History)/Natural Environment Research Council Study Group for a co-ordinated display of minerals, gemstones, rocks and meteorites in the BM(NH) and the Geological Museum.

These proposals will necessarily involve a considerable upheaval in the public galleries during the next decade. In order to minimize the effects of this disruption: (1) A temporary display will be provided when any major group of organisms is not represented in a permanent display for a substantial period of time. (4) Access to the Whale Hall will be maintained even though work may be in progress there; thus the Blue Whale will be continuously on display to the public. (3) The larger fossil mammal specimens will be inaccessible only during the period of transit from one gallery to another, as was the case with dinosaurs. And (4), the insects will be off display for a matter of weeks only.

Changes of this magnitude will cause many difficulties and the museum welcomes constructive suggestions, either now or as the work proceeds. The trustees are aware of the controversy aroused by the first phases of the new policy. However, a paper by Dr G. C. S. Clarke and Dr R. S. Miles² of the museum's Department of Public Services explained the ideas behind the changes which had provoked the controversy and invited constructive criticism. No replies were received. The museum has continuously monitored the effectiveness of the new exhibition and the responses of the visiting public and has good evidence that the vast majority of the visiting public, especially schoolchildren and their teachers, greatly appreciate the new exhibitions³. Suitable modifications have been made when they would lead to improvements.

To assist visitors to understand the complicated changes which are taking place the museum is preparing a small temporary exhibit of the plans.

T. R. E. SOUTHWOOD
(Chairman of the Board of Trustees)
R. H. HEDLEY
(Director)

British Museum (Natural History),
London SW7, UK

1. *Report of the British Museum (Natural History), 1972-74*, p.3 (Trustees of the British Museum (Natural History), 1975).
2. Clarke, G. C. S. & Miles, R. S. *Biologist* 27, 81-83 (1980).
3. Alt, M. B. *Mus. J.* 80, 10-19 (1980); also unpublished reports on the school use of the Natural History Museum, 1979 and 1980.

Exhibitions planned in the next ten years

Subject	Location
<i>Origin of Species</i> (to be opened in May 1981)	Upper South-West Gallery
<i>Mammal diversity</i> (1) — cetaceans, artiodactyls, perissodactyls, hyraxes, sirenians, elephants, rodents and related fossil orders	Whale Hall
<i>Mammal diversity</i> (2) — mammal-like reptiles, monotremes, marsupials, edentates, carnivores, insectivores, bats and fossil groups	Galleries now used for insects and marine invertebrates
<i>Mammal diversity</i> (3) — primates	West balcony of Central Hall
<i>Arthropod diversity</i> (1) — insects centipedes, millipedes and Onychophora	East Pavilion and South-East Gallery
<i>Arthropod diversity</i> (2) — Crustaceans	South-East Gallery
<i>Arthropod diversity</i> (3) — spiders, scorpions, ticks, mites and trilobites	South-East Gallery
<i>Unity in diversity</i> — an introduction to all the museum's exhibitions	Entrance to South-East Gallery

NEWS AND VIEWS

Spatial frequency analysis in vision

from Oliver Braddick

READERS of *Nature* will probably be aware that many experimenters and theorists in visual perception and physiology have taken to using the language of Fourier analysis: they test vision with sine wave grating patterns and discuss the behaviour of the visual system in terms of spatial frequency response. Why has this approach become popular, and has it been justified as enhancing our understanding of vision?

Descriptions in terms of frequency are most familiar when speaking of periodic events in time, such as sound or radio waves, but they may be used for spatial patterns in an analogous way. The number of repetitions of a pattern within a unit of distance is its spatial frequency. The mathematics of Fourier analysis which allow any complex sound to be described as a set of harmonics also allow any visual pattern to be described as a sum of spatial frequency components, that is, patterns whose luminance varies across space as a sine wave. A system designed to transmit time-varying signals, such as a hi-fi set, commonly has its performance specified by its frequency response. Similarly, how well a system transmits spatial patterns can be characterised by its response to different spatial frequencies. In optics, this has been a commonplace technique for evaluating lens systems for many years. The performance of the human eye, however, is determined not only by its optical quality in forming the retinal image, but also by how well the visual system signals the information in that image. It therefore seemed natural to extend the idea of measuring spatial frequency response to the whole system^{1,2}. This can be done if we make the fairly uncontroversial assumption that for a human subject trying to detect a low-contrast grating pattern, a low threshold indicates a high response of the visual system to that spatial frequency, and so on, using thresholds to measure the contrast sensitivity function.

The spatial frequency response of some component within the visual system may also be measurable: this can be done directly in physiological experiments by measuring the response of single nerve cells to gratings of various frequencies, and

indirectly in psychophysical experiments by testing whether two gratings interact in their effects on perception. For example, if one grating can mask detection of another, it is concluded that they must be processed by a common neural 'channel'³.

The flood of experimental work in what has been dubbed 'gratingology' has come largely from the idea proposed by Campbell and Robson⁴ that Fourier methods were valuable, not just for characterising the performance of the whole visual system, but also for understanding the way the system is divided into parallel information-carrying channels. It is generally believed that the properties of these channels define the 'language' in which visual information is conveyed to, and perhaps within, the brain. The idea that this language is best analysed in terms of spatial frequency has been vigorously asserted and contested. If we are to avoid this controversy obscuring the significance of the experimental results, we must keep clear the distinctions between various levels at which spatial frequency methods and Fourier theory may be applied to vision.

First, the spatial frequency response of the visual system can be used as a valuable descriptive measure, without commitment to any particular assumptions about the properties of the visual system. Second, insofar as the visual system, or some component of it, can be treated as linear, the spatial frequency response has important predictive power: since any pattern can be considered as the sum of sine wave components, the visual response to any pattern can be predicted by summing the responses to the separate components. Third, it has been proposed that the visual system at some level consists of parallel, independent channels, each responding to its own limited band of spatial frequency. If so, the visual system itself is performing, at least crudely, an analysis of the visual pattern into its frequency components. Fourth, and most radically, it is possible that this division of the visual pathway into

spatial frequency channels might indicate a fundamental way in which visual information is organized: that is, perceptual processes such as selective attention, or recognition, might operate on representations of the visual world in terms of spatial frequency components rather than the more familiar kind of representation in terms of spatial location.

The descriptive use of spatial frequency analysis, at its simplest, amounts to saying that it is important to know not only how well vision conveys very fine detail (high spatial frequencies), but also how well it conveys information about the patterning of light and dark in the image at every scale from the very coarse to the very fine. One application of this is clinical: the quality of vision is conventionally measured in terms of acuity (that is, detail resolution) but some patients complain of poor vision even when their measured acuity is normal. In many cases they can be shown to have abnormally low sensitivity to spatial frequencies below the acuity limit. Indeed, differences in the pattern of contrast sensitivity may be diagnostically significant⁵. An important side effect of the descriptive use of spatial frequency response has been to emphasise the usefulness of contrast as a parameter of vision. Even for non-sine wave stimuli, contrast threshold is often a more useful measure than luminance threshold, since contrast can be varied without introducing the radical changes that occur in the visual system with adaptation to different light levels.

If spatial frequency response is found to have predictive as well as descriptive power, that implies that the visual process concerned shows linear spatial summation. This in turn implies that it could be described equally completely in spatial as in spatial frequency terms. This is true, for instance, of one class of retinal ganglion cells (X-cells)⁶: their response to a pattern can be predicted either from the summation of response to its spatial frequency components, or from summation of the local responses given by each point in the receptive field. The spatial sensitivity profile of the receptive field and the spatial frequency response function may be

Oliver Braddick is in the Department of Experimental Psychology, University of Cambridge.

mathematically transformed into each other. In experimental practice, this complete description may be more readily achievable by testing the spatial frequency response, since the cell's response to a grating which covers the whole receptive field is stronger and clearer than that to illumination of a small spot at some point within it.

It can be argued, though, that the spatial frequency approach has a particular value when the requirement of linearity is *not* met. This is because frequency analysis provides tests that very clearly indicate the presence and nature of any non-linearity present. Thus, it was through the investigation of response to sinusoidal stimuli that an important class of non-linear (Y) retinal ganglion cells was discovered by Enroth-Cugell and Robson⁶, and that the nature of the mechanisms which produce their characteristic behaviour has begun to be understood⁷.

Taking the performance of the visual system as a whole, there are clearly various significant non-linearities, but for the conditions under which many detection experiments are done linear predictions work quite well⁴. This means that it is useful to consider the separate visual response to each of the frequency components making up a complex pattern. Using this mathematical analysis does not require us to suppose that the visual system breaks up the pattern into these components in any physical sense. The frequency response can be taken to describe the properties of visual system as a homogeneous linear spatial filter — the 'single channel' model. The alternative is a 'multi-channel' model in which the analysis into spatial frequency components has more physical reality. In this model, the overall frequency response is the envelope of the sensitivity functions of many independent channels, and different channels may respond to components in a complex pattern. A wide range of psychophysical experiments support the idea of multiple spatial-frequency channels^{3,8}, although there are still considerable quantitative problems to be resolved on how many parallel channels there are⁹, and how narrow is the band of spatial frequencies to which each responds¹⁰.

Physiological experiments, too, show that the visual cortex contains cells responding to different bands of spatial frequencies in the same region of the visual field¹¹. It is true that, insofar as these cells show linear spatial summation (as many of them do), their receptive field sensitivity profiles equally describe their properties¹². For fields of similar form, the differences in optimal frequency will correspond to differences in receptive field size. Even so, the frequency description has a good deal of heuristic value. For instance, a qualitative account in terms of receptive field dimension might lead one to expect that such cells would differ strikingly in the width of bright (or dark) bar to which they

respond best. In fact, this turns out to characterise their differences much less clearly than does their spatial frequency response¹³.

However, spatial frequency response alone does not give all the information expressed in a spatial sensitivity profile. This is because the response is a function of the phase as well as the spatial frequency of a grating. A cell which responds well to bar patterns will show optimal response to different grating frequencies when they have a peak of the waveform in a given position, while responsiveness to edge patterns corresponds to the requirement that waveforms of different frequency should have their maximum gradient in the same position. What is apparent from making the distinction in terms of phase, rather than edges versus bars, is that any pattern can be considered a mixture of the two phase relationships. Both psychophysicists and physiologists¹² are now exploring the possibility that the division into two types of phase response is another form of 'channeling' for encoding spatial patterns in vision.

Given the existence of channels that respond independently to different spatial frequency bands, should we conclude that visual perception operates in the Fourier domain? In modern digital image processing, it is often found useful to perform various operations on an array of, say 256×256 spatial frequencies, each of which is derived in the Fourier transform of an entire picture of 256×256 local elements. It is most unlikely that visual perception works in anything like this way. Physiological and psychophysical evidence

concur that, at the levels we have been considering, analysis is fairly local: receptive fields are limited in size. There is a mathematical inverse relationship between the width of the spatial frequency band to which a channel responds and the range of spatial positions to which it responds¹⁴. It appears that the visual system represents a compromise, dividing the spatial frequency spectrum fairly coarsely within local, but not very small, areas.

In fact, it would probably be pointless for the visual system to measure spatial frequency components, or anything else, over the visual field as a whole. What we usually want to recognise is not the visual field as a unified entity, but objects within it. Could we be recognising an object by means of the spatial frequency components detected by a local set of channels? In most cases, this seems quantitatively unlikely. Analyses of spatial frequency channels in central vision have generally found them only for spatial frequencies above 1 cycle/degree. Such channels could encode an adequate frequency spectrum only for an object that was roughly $\frac{1}{2}$ degree across, or smaller. With the important exception of written characters, we need to recognise most objects over a range of visual angles markedly larger than this. A similar conclusion arises from physiological receptive field sizes, and from the extent of visual field analysed by a single cortical 'hypercolumn'¹⁵: at the level of visual cortex, a single object will normally be analysed by many, separate, receptive fields. An important aspect of the processing of visual shape must be the 'grouping' of information from local analysers spread out over an area of the visual field.

What, then, might be the role of spatial frequency analysis in the process of perception? Marr and Hildreth¹⁶ have suggested that coincident signals in different frequency channels provide an effective criterion for finding edges in the visual field. In this view spatial frequency analysis is used at a rather low level to define features, which themselves are part of a representation in the domain of space rather than spatial frequency. Separate processing of spatial frequency bands has been proposed as playing a rather similar role of local disambiguation in the computation of stereo disparity^{17,18}. In these models, the information signalled by spatial frequency channels is not separately available for higher perceptual processes. This argument has been supported by experimental evidence that patterns can differ quite markedly in the Fourier domain without it being possible to use these differences for perceptual grouping¹⁹ (although this result concerns the ability to separate components that differ in orientation, not in spatial frequency itself).

On the other hand, one of the best known effects in grating psychophysics is the 'perceived spatial frequency shift' produced by adaptation to a grating²⁰. This

1. Schade, O.H. *J. opt. Soc. Am.* **46**, 721 (1956).
2. Campbell, F.W. & Green, D.G. *J. Physiol., Lond.* **181**, 576 (1965).
3. Braddick, O., Campbell, F.W. & Atkinson, J. in *Handbook of Sensory Physiology, Vol. VIII: Perception* (eds Held, R., Leibowitz, H. & Teuber, H.L.) (Springer, Heidelberg, 1978).
4. Campbell, F.W. & Robson, J.G. *J. Physiol., Lond.* **197**, 551 (1968).
5. Arden, G.B. *Br. J. Ophthalmol.* **62**, 198 (1978). Bodis-Wollner, I. & Diamond, S.P. *Brain* **99**, 695 (1976). Hess, R.F. & Howell, E.R. *Vision Res.* **17**, 1049 (1977).
6. Enroth-Cugell, C. & Robson, J.G. *J. Physiol., Lond.* **187**, 517 (1966).
7. Hochstein, S. & Shapley, R.M. *J. Physiol., Lond.* **262**, 237 (1976).
8. Blakemore, C. & Campbell, F.W. *J. Physiol., Lond.* **203**, 237 (1969). Graham, N. & Nachmias, J. *Vision Res.* **11**, 251 (1971). Graham, N., Robson, J.G. & Nachmias, J. *Vision Res.* **18**, 815 (1978). Sachs, M., Nachmias, J. & Robson, J.G. *J. opt. Soc. Am.* **61**, 1176 (1971).
9. Wilson, H.R. & Bergen, J.R. *Vision Res.* **19**, 19 (1979). Watson, A.B. *Vision Res.* (in the press).
10. Graham, N. *Vision Res.* **17**, 637 (1977). King-Smith, P.E. & Kulikowski, J.J. *Vision Res.* **21**, 235 (1981).
11. Maffei, L. & Fiorentini, A. *Vision Res.* **13**, 1225 (1973).
12. Movshon, J.A., Thompson, I.D. & Tolhurst, D.J. *J. Physiol., Lond.* **283**, 53 (1978).
13. De Valois, R.L., Albrecht, D.G. & Thorell, L.G. in *Frontiers in Visual Science* (eds Cool, S.J. & Smith, E.L. III) (Springer, New York, 1977).
14. Mackay, D.M. *Nature* **289**, 117 (1981).
15. Hubel, D.H. & Wiesel, T.N. *Proc. R. Soc. Lond.* **B198** (1977).
16. Marr, D. & Hildreth, E. *Proc. R. Soc. Lond.* **B207**, 187 (1980).
17. Marr, D. & Poggio, T. *Proc. R. Soc. Lond.* **B204**, 301 (1979).
18. Mayhew, J.W.W. & Frisby, J. *Perception* **9**, 60 (1980).
19. Mayhew, J.E.W. & Frisby, J.P. *Nature* **275**, 438 (1978).
20. Blakemore, C., Sutton, P. & Nachmias, J. *J. Physiol., Lond.* **210**, 727 (1970).
21. Campbell, F.W. & Marr, D. Discussion in *Phil. Trans. R. Soc.* **B209**, 9 (1980).

effect is best explained by the assumption that perceived spacing of the grating is directly determined by which spatial frequency channel responds most strongly. It can be argued that judging the spatial frequency of a grating is more related to recognising a repetitive texture, in which we are unconcerned with the spatial identity of individual features, than to perceiving the size and shape of a discrete object. Maybe objects are represented in spatial terms, but the textures of their surfaces in spatial frequency terms.

Speaking of 'objects' should not lead us into the assumption that there we divide any particular visual scene in a fixed way. Clearly for some purposes we are primarily concerned with the gross layout of our visual environment, while for others we need information that is present in fine detail. It is tempting to suppose that for different perceptual tasks we might switch our attention to different spatial frequency bands. However, it is hard to find an

experimental example that demands an explanation as spatial frequency tuning of attention, and there are certainly instances where we do *not* find it possible to select a frequency band in this way²¹.

Many visual scientists who are sceptical or open minded on how far higher perceptual processes speak the language of spatial frequency nonetheless find that language invaluable for describing and manipulating their experimental stimuli and the information signalled in the visual system. New insights come from thinking of the visual image as containing patterns of light and dark over a range of spatial scales, or 'degrees of blurredness'. Formally, this idea, and analyses that embody it, may be expressible as exactly in spatial as in spatial frequency terms. The fact remains that by exploring the implications of the spatial frequency description, and by using stimulus patterns which were simple to describe that way, we have learned a great deal about vision and by all indications have more to learn yet. □

increasing frequency until it coincides with the high field d.c. value at ~ 100 MHz (refs 7 and 9). It soon became clear that the electric field is not destroying the charge density wave, because an X-ray diffraction experiment shows that the superlattice spots are not affected by a current through the sample even when the current exceeds the linear threshold¹⁰. Diffraction experiments reveal incommensurate lattice distortions with periods λ such that $a/\lambda = 0.24$ and 0.26 for the upper and lower transitions. These data are consistent with an interpretation in terms of impurity pinning of the charge density wave¹¹. The charge density wave can undergo small oscillation, thus accounting for the frequency-dependent conductivity. The pinning frequency is very low, possibly because the impurity concentration in NbSe₃ is very small. What makes NbSe₃ special among charge density wave systems in this respect is not entirely understood at present. The weakness of the pinning also manifests itself in the low-frequency dielectric constant, which has been measured to be greater than 10^8 (see ref. 9). Presumably a small electric field is sufficient to overcome the weak pinning so that the charge density wave can slide in the way envisioned by Fröhlich — an interpretation supported by an experiment in which impurities were introduced into NbSe₃ and the threshold electric field was found to increase¹². Instead of classical depinning, an alternative mechanism involving the quantum mechanical tunneling of macroscopic segments (up to several microns in length) of charge density waves has been proposed¹³, but all existing theories are based on the collective motion of the charge density wave ground state described by Fröhlich.

Further study of NbSe₃ yielded more surprises. When the current exceeds the threshold, voltage noise across the sample greatly increases. Superimposed on this background noise are periodic components with well defined frequencies⁸, which scale linearly with the excess nonlinear current¹⁴ and are typically in the range of kHz to MHz. A model for the noise is that it arises when the charge density wave moves through the impurities or the lattice, much like a particle rolling down a washboard. If certain assumptions are made about the charge being carried by the sliding charge

Sliding charge density waves

from P. A. Lee

THE idea of a charge density wave transition goes back to Peierls¹ and, independently, to Fröhlich² in 1954. They reasoned that a one-dimensional metal can lower its energy by distorting the lattice and forming a gap at the Fermi surface, thereby making a transition into an insulating state. The term 'charge density wave' refers to the fact that the lattice and the electron charge density form a new periodic structure, with a wavelength λ that is longer than the original lattice period a . Fröhlich further reasoned that if the charge density wave is incommensurate, that is, a/λ is not a simple rational fraction like $1/2$ or $2/3$, then the entire charge density wave structure can slide through the lattice, thereby contributing to the electrical conductivity. Since these ideas were originally formulated in terms of one-dimensional systems, for many years they were considered with nothing more than theoretical curiosity.

In the early 1970s the interests of many physicists turned to quasi one- and two-dimensional systems. The existence of a charge density wave ground state in these systems turns out to be the rule rather than the exception, and a large number of examples have been discovered. Experimentally the onset of the charge density wave is signaled by a rise in the resistivity at some temperature T_c . In quasi one-dimensional systems a transition to an insulating state often occurs, whereas in the layered compounds the system often

stays metallic. The most direct microscopic evidence for a charge density wave transition is the appearance of new superlattice spots associated with the new lattice periodicity λ which is detected by X-ray, neutron or electron diffraction. Prominent examples of charge density wave systems are the organic charge transfer salts³, such as tetrathiafulvalene-tetracyanoquinodimethane (TTF-TCNQ) and layered compounds⁴ formed from transition metals and chalcogens, such as TaS₂ and NbSe₃.

However, until 1976, none of the charge density wave systems had exhibited the sliding conductivity envisioned by Fröhlich. This is because the charge density wave was either commensurate or pinned by impurities. Impurities are sensitive to the charge density oscillations and can lock the charge density wave in place, a phenomenon very analogous to static friction⁵. As in friction, a sufficiently large force should dislodge the charge density wave. Thus the observation^{6,7} in 1976 of nonlinear electrical conductivity in NbSe₃ was greeted with a good deal of interest. In NbSe₃ there are two transitions at 59K and 144K; below each of these temperatures the resistivity rises. It was found that beyond a certain threshold electric field⁸, which can be as low as a few meV per cm for the low-temperature transition and 0.1 eV per cm for the higher-temperature transition, the resistivity becomes nonlinear and begins to drop, until it saturates to a value close to that expected if the charge density wave transition had not occurred.

The conductivity is also found to be highly frequency dependent, rising with

1. Peierls, R. E. *Quantum Theory of Solids* (Oxford University Press, London, 1955).
2. Fröhlich, H. *Proc. R. Soc. A* **223**, 296 (1954).
3. *The Physics and Chemistry of Low Dimensional Solids* (ed. Alcaicer, L., 1979).
4. Wilson, J. A. *et al. Adv. Phys.* **24**, 117 (1975).
5. Lee, P. A. *et al. Solid State Commun.* **14**, 703 (1974).
6. Monceau, P. *et al. Phys. Rev. Lett.* **37**, 602 (1976).
7. Ong, N. P. & Monceau, P. *Phys. Rev. B* **16**, 3443 (1977).
8. Fleming, R. M. & Grimes, C. C. *Phys. Rev. Lett.* **42**, 1423 (1979).
9. Grüner, G. *et al. Phys. Rev. Lett.* **45**, 935 (1980).
10. Fleming, R. M. *et al. Phys. Rev. B* **18**, 5560 (1978).
11. Lee, P. A. & Rice, T. M. *Phys. Rev. B* **19**, 3970 (1979).
12. Ong, N. P. *et al. Phys. Rev. Lett.* **42**, 811 (1979).
13. Bardeen, J. *Phys. Rev. Lett.* **45**, 1978 (1980).
14. Monceau, P. *et al. Phys. Rev. Lett.* **45**, 43 (1980).
15. Fung, K. K. & Steeds, J. W. *Phys. Rev. Lett.* **45**, 1696 (1980).

P.A. Lee is at the Bell Laboratories, Murray Hill, New Jersey.

density wave (a quantity under debate at present), the frequency of the noise can be accounted for¹⁴.

Recently there was a report of dark-field electron microscopy on NbSe₃ (ref. 15). An image of an area of the sample is formed using the superlattice reflection alone; thus any feature in the imaging is due to the charge density wave formation. It is found that stripes approximately 200 Å wide run along the conducting axis, and across the stripes are bands of lighter and darker intensities, suggestive of a Moiré pattern. It appears that the charge density wave structure may itself be broken into domains and that more than one superlattice wave vector is possible, so that interference between them forms the Moiré

pattern. What is even more intriguing is that at low temperatures the bands are found to shift in time, so that the entire pattern shimmers on a time scale of a fraction of a second. This suggests some time-dependent motion of the charge density wave, but it is occurring on such a slow time scale that its connection with the nonlinear and frequency-dependent conductivity is unclear.

It is rare to find a material like NbSe₃ that exhibits such a rich variety of unusual properties. At present the idea of a sliding charge density wave provides a framework for the qualitative understanding of many of these properties, but clearly much work remains to be done before a quantitative understanding is achieved. □

insertion was shown by E.S. Belyaeva[†].

The localization of the mdg elements in chromosomes appears highly conserved during differentiation as they have been found in the same position in cells of salivary glands and stomach (E.V. Ananiev[†]). They are also stable in inbred lines of *D. melanogaster*. The probability of 'jumping' for mdg1 and mdg3 was estimated to be of the order of 10⁻⁵.

Drastic changes in mdg1 and mdg3 localization were found to take place during back selection for high sexual activity among inbred flies characterized by low sexual activity (V.S. Gvozdev[†]). Such transpositions might facilitate adaptation of the organism to the environment through influence on the expression of other genes.

Mdg elements appear species specific: those of *D. melanogaster* cannot be detected in *D. virilis* and vice versa (M.B. Evgen'ev^{*}). A new mdg family has been cloned by G. N. Yenikolopov from *D. virilis* genome. *D. littoralis* chromosomes included in the genome of *D. virilis* are able to accept *D. virilis* mdg copies without crossing-over.

A.N. Gudkov (Oncological Centre, Moscow) discussed endogenous proviruses and their similarity to mdg elements. Different strains of hen differ strikingly in the number and localization of proviral sequences as studied by Southern blotting analysis. The existence of a natural strain completely lacking RAVo-like sequences was described. Together with the data of Astrin *et al.* (*Cold Spring Harb. Symp. quant. Biol.* 344; 1105, 1979) this suggests that the presence of proviral sequences is not essential for the organism.

Sequences homologous to the oncogene of mouse Moloney sarcoma virus (putative protooncogene) and to some other sequences of the virus (human C-type provirus) are found in the normal human genome (E.A. Zabarovsky^{*}). The protooncogene-containing sequence has been cloned from the human genome.

D.M. Kramerov^{*} described the properties of two mdg-like elements cloned from the mouse genome selected on the basis of efficient hybridization to long stretches of double-stranded RNA (dsRNA-A). One of these cloned sequences was found to be responsible for the synthesis of intracisternal A-particles. Also, endogenous C-type proviruses were detected among dsRNA-A-hybridizing sequences of mouse genome.

Structural similarities between mdg elements of *Drosophila* and endogenous proviruses were stressed by Georgiev. He also discussed a 'promoter' hypothesis of carcinogenesis (see *J. theor. Biol.* 25;473, 1969) which considers oncogenic transformation to result from a potential cellular oncogene falling under the control of the viral promoter. Mdg elements may provide an important source of such poorly controlled promoters in the course of non-viral oncogenesis. □

Mobile genetic elements: new Soviet studies

from A. D. Mirzabekov

A CONFERENCE devoted to studies of eukaryotic mobile dispersed genetic elements was held on November 14, 1980 as a session of the seminar 'Chromosoma' at the Institute of Molecular Biology (USSR Academy of Sciences, Moscow).

In the *Drosophila melanogaster* genome mobile dispersed genetic (mdg) elements are multiple, scattered throughout the whole genome, and can move around. They are also actively transcribed (G.P. Georgiev^{*}). At present, about 20 families of mdg elements have been described and comprise as much as five per cent of the whole *D. melanogaster* genome. The variable location of mdg elements is the most striking example known of genomic differences between individuals of the same species.

Y.V. Ilyin[†] described the structural organization of the two elements, mdg1 (a ~7 kilobase (kb) element including the previously described Dm225 and Dm234 sequences) and mdg3 (~5 kb) cloned in the plasmid pBR322. In cell culture, mdg1 and mdg3 are amplified but not tandemly repeated. Like all other transposons, mdg1 and mdg3 contain long terminal repeats (LTRs) at the ends of the gene.

Sequence data show the structure of the LTRs of mdg3 (Bayev *et al.* *Nucleic Acids Res.* 8; 3263, 1980; A.S. Krayev, V. Kulguskin^{*}) to be perfect direct repeats 268 bp long flanked by two mismatched 18 bp inverted repeats. The complete mdg 3 element is then bounded by short (5 bp) direct repeats probably derived from host target sequence duplication. Mdg1 contains longer perfect direct repeats (442 bp) flanked by mismatched 16 bp inverted

repeats. In this respect, the organization of mdg1 LTR is similar to LTRs of the *copia* gene (Levis *et al.* *Cell* 21; 580, 1980) or terminal repeats of integrated proviruses (Beveren *et al.* *Proc. natn. Acad. Sci. U.S.A.* 77; 3307, 1980; Dhar *et al.* *ibid*; 3937, 1980). Within both LTRs, signal sequences for transcription initiation (Hogness box) and polyadenylation (AATAAA) are found.

The whole mdg sequences of mdg1 and mdg3 are transcribed, initiation and termination sites being localized within LTRs (V.G. Chmelauskaite^{*}). Both DNA strands are transcribed giving rise to the long double-stranded RNA stretches found in total poly(A)⁺ RNA from cell culture. However, the efficiency of transcription of one strand is much higher than that of the other. At least in *D. melanogaster*, hybridization to very long double-stranded RNA can be used as a diagnostic test in the search for new mdg elements. Besides full-length transcripts, spliced poly(A)⁺ RNAs were found among mdg1- and mdg3-specific RNA sequences. The major part of mdg transcripts are located in free ribonucleoprotein particles in the cytoplasm and only a minor fraction is bound to polysomes.

M.D. Golubovsky (Institute of Cytology and Genetics, Novosibirsk) showed that the frequency of an unstable mutation in the *sn* locus of *D. melanogaster* can simultaneously increase in geographically distant natural populations of flies. This may be due to 'infection' with a movable element. In addition, a mutation was found which influences both *sn* and *ct* loci and probably depends on the presence of the same element. A correlation between instability in *white* locus and mdg 1

^{*} Institute of Molecular Biology, Moscow.

[†] Institute of Molecular Genetics, Moscow.

Darwin's missing notebooks come to light

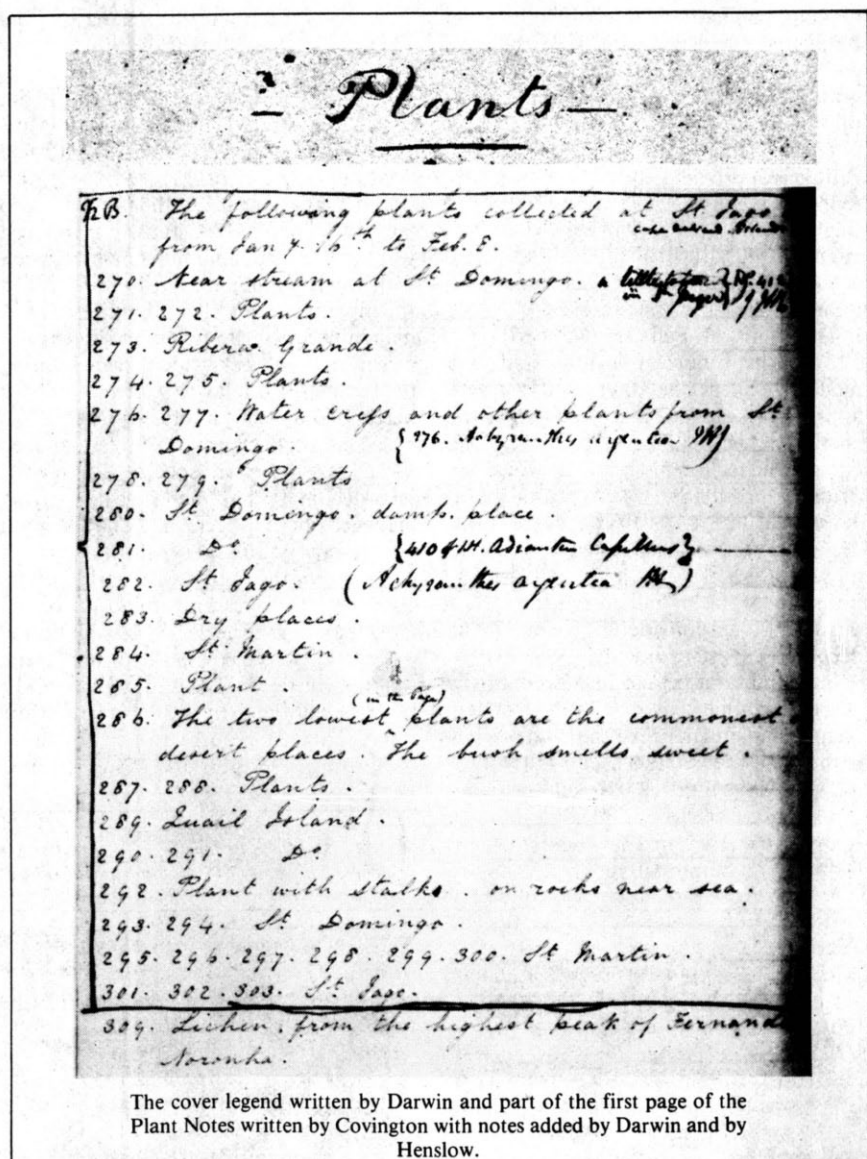
from Duncan M. Porter

IN early December 1980, Rita I'Ons, Assistant at the Cambridge University Herbarium, discovered two notebooks titled *C. Darwin's Plants from S. America* and *Darwin's S. American Plants* in a cupboard at the Herbarium. The notebooks, on the plants collected during the voyage of HMS *Beagle* (1831–1836) had been lost for more than a hundred years. Their long absence has caused many problems with the identification of Darwin's plants and has also prevented the botanical aspects of his voyage from being properly appreciated.

The notebooks are in the writing of J.S. Henslow, Professor of Botany at Cambridge and Darwin's mentor, who originally agreed to identify the plants. However, because of other commitments, Henslow was unable to identify more than a few, and the task fell to J.D. Hooker at Kew in the early 1840s. Henslow sent Darwin's notes to Hooker, and presumably they were later returned, but earlier searches at Cambridge and Kew failed to find them¹. Hooker and others identified a number of Darwin's plants, but the majority remain unclassified to this day. Tied with string into the back of the second notebook are 11 pages of notes titled 'Plants'. They are in the handwriting of Syms Covington, Darwin's servant and amanuensis on the *Beagle* and until 1839. A few notes and corrections have been added in Darwin's hand, as have a few notes by Henslow.

Notes on the plants vary in length from the terse 'Plant' to a page-long discussion on the Wild Potato of Chile's Chonos Archipelago (*Solanum tuberosum* var. *guaytecarrum*). The latter was used as the basis for the discussion in Darwin's *Journal of Researches*². Likewise, one can trace other notes to passages in the *Journal*. Some, such as that on the Dwarf Beech of Tierra del Fuego (*Nothofagus antarctica*), make more interesting reading than when published: "Jemmy Button said. 'when leaves yellow, snow all go'. — Captain FitzRoy states that in April the leaves of the trees which grow on the lower parts of the hills turn colour; but not those high up. — I recollect having read a paper to show that in England warm Autumns hastened the falling of the leaves; that the process is a regular part of the vegetation: This fact would seem to show the same law." When published in the *Journal*, Button's quote was excised.

A passage in the Plant Notes describing the forests of Tierra del Fuego seems to anticipate the 'tangled bank' of the *Origin*



The cover legend written by Darwin and part of the first page of the Plant Notes written by Covington with notes added by Darwin and by Henslow.

of *Species*: "The extreme dampness of the climate favours the course luxuriance of the vegetation: the woods are an entangled mass where the dead and the living strive for mastery." In the last paragraph of the first edition of the *Origin*³, Darwin writes: "It is interesting to contemplate an entangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about, and with worms crawling through the damp earth, and to reflect that these elaborately constructed forms, so different from each other, and dependent upon each

other in so complex a manner, have all been produced by laws acting around us."

The Plant Notes follow the same format as the lists of notes on *Beagle* zoological specimens at the Cambridge University Library and the British Museum (Natural History). Each series of notes was compiled for the taxonomists who were to identify the various groups of organisms collected (plants, birds, fish, insects, mammals, molluscs, and reptiles and amphibians). As Darwin intended, the Plant Notes are proving quite useful in the identification of his *Beagle* specimens. □

Duncan M. Porter is Associate Professor of Botany at Virginia Polytechnic Institute and State University, Blacksburg, and Visiting Fellow at Clare Hall, University of Cambridge.

- Porter, D. M. J. *J. Soc. Bibliophy nat. Hist.* 9, 515 (1980); *Bot. J. Linn. Soc., Lond.* 81, 79 (1980).
- Darwin, C. *Narrative of the Surveying Voyages of His Majesty's Ships Adventure and Beagle, Between the Years 1826 and 1836, Describing Their Examination of the Southern Shores of South America, and the Beagle's Circumnavigation of the Globe*, Vol. 3, *Journal and Remarks, 1832-1836* (Colburn, London, 1839); second

- edn: *Journal of Researches into the Natural History and Geology of the Countries Visited During the Voyage of H.M.S. Beagle Round the World, Under the Command of Captain FitzRoy, R. N.* (Murray, London, 1845).
- Darwin, C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, 489 (Murray, London, 1859).

The well-managed lawn

from Peter D. Moore

ONE of the most widespread vegetation types, and one which is certainly the most easily accessible to urban populations, is that highly managed piece of anthropogenic grassland, the domestic lawn. It has been estimated that the area of lawn in Britain is approximately 2,700 km² (NERC report, *Amenity Grasslands*; 1977), plus a further 1,100 km² of playing fields. It is surprising, therefore, that so little academic research has been conducted upon so amenable a habitat.

The work of Falk (*Ecology* 57; 141, 1976) on his suburban lawn in California made it evident that the lawn is a very appropriate ecosystem for studies of the energetic type. Falk produced a very full energy budget for his lawn, in which net primary productivity was estimated at 4,468 kcal m⁻² yr⁻¹ (1.05 kgm⁻²). Of this 10.4 per cent was harvested by man as mowings, 2.9 per cent was consumed by invertebrate grazers and the remaining 86.7 per cent was considered to enter the decomposer system directly.

Falk's data also take into account the energy investment by man in the management and maintenance of the lawn, mainly in the form of fossil fuel used by the motor mower. This totals 2,029 kcal m⁻² yr⁻¹. Thus if we were to consider the lawn in terms of harvest efficiency the outcome is low, with an output : input energy ratio of

0.22. The ratio obtained (Blaxter, *Nature* 252; 531 1974) for the total energy efficiency of British agriculture is 0.4.

The British lawn is still neglected, but Falk has now provided us with more information about the lawns of the United States (*J. appl. Ecol.* 17, 689; 1980), this time in the east, near Washington DC. Two lawns were selected for study, one received what was termed 'minimum maintenance', with no irrigation or fertilization and a fortnightly mowing. The other site was well maintained, being fertilized and limed once a year, mowed weekly and irrigated during the summer. Both lawns were mowed to a height of 5 cm using a rotary mower.

As one might expect, the undermanaged lawn was richer in species (22 species recorded in 0.7 ha) than was the highly managed one (11 species in 0.2 ha). What is surprising is the very similar values obtained for the total net primary productivity (below-ground production is included). For the neglected lawn this came to 1,669 kg m⁻² yr⁻¹ (6,380 kcal) and for the well-managed lawn 1,649 kg m⁻² yr⁻¹ (6,814 kcal). So all the extra management effort serves only to reduce diversity without increasing productivity. The amount

removed in mowings is about 20 per cent of the total productivity, rather higher than that observed in California.

As Falk points out, the mowed lawn in temperate climes is a highly productive ecosystem, despite its low stature. The highest productivity of tropical grasslands (around 2,000 g m⁻² yr⁻¹) and the productivity of maize or wheat in temperate environments (1,000–1,600 g m⁻² yr⁻¹) are comparable to that of the suburban lawn. It is salutary to reflect, therefore, upon the energy harvest each year from such habitats in Britain alone. If these can match the productivity of their Washington counterparts (and anyone who has seen the lawns of Washington in August will have no hesitation in allowing this), then 332 kg m⁻² (1,320 kcal m⁻²) is harvested as mowings each year from our 3,800 km² of urban and suburban amenity grasslands. This amounts to 1.2 × 10⁹ tons of dry matter, or 5 × 10¹² kcal each year. Taking the original estimate of energy output : input ratio of 0.22, this implies that we expend a total of 2.27 × 10¹³ kcal on the harvesting of this herbage each year in Britain. Given the high energy investment in maintaining the British lawns and cricket pitches, the energy-from-biomass lobby must be tearing out its hair to see mowings being fed to the decomposer systems of the nations's compost heaps. □

Peter D. Moore is in the Department of Plant Sciences, University of London King's College.



100 years ago

THE GREAT VIENNA TELESCOPE

THE political and social disturbances in Ireland have of late somewhat diverted attention from the literary and scientific work done in that country. Such work has nevertheless proceeded on its quiet way despite land agitation, failure of crops, or even commercial distress; and Ireland is to be congratulated on the completion of the fine 27-inch refracting telescope, designed and constructed by Mr. Howard Grubb of Dublin for the Imperial and Royal Observatory of Vienna. This telescope is the largest equatorial refractor at present in existence.

In the largest telescope until now in existence, the great 26-inch refractor at Washington, the oft-recurring operation of reading the circle involved the sending of an assistant with a hand-lamp to climb up some twenty feet of the instrument, and the vernier not being in a fixed place, but moving about with the telescope, the labour and difficulty of this operation are great. In the Vienna instrument Mr. Grubb has so contrived it that the observer sitting in his chair can read the right ascension and declination circles through one single reader, all being illuminated by one gas-lamp hung on the end of the declination axis.

Another wonderful convenience is enabling an assistant to easily turn and set the instrument to right ascension. This can be understood from the figure,



which is from a photograph. In the case of an instrument of the size of the Vienna equatorial, the observer requires an elaborate stage or platform of variable height and position in order to reach the eyepiece, and if he has to move to another object he must descend from his chair or platform and move it before he can move the instrument, and then he has not the facility of sighting objects, but must adopt a sort of tentative process, climbing up and down his stage and moving it and the telescope alternately. To avoid this labour and save time an arrangement is supplied by

which the assistant from the ground floor can set the instrument in right ascension, and once set in one direction, the other (declination setting) can be readily managed by the observer.

We learn that M. Alphaud, the Director of Public Works in Paris, has in his hands the tender of Siemens and Co. for constructing an electrical railway from the Place de la Concorde to the interior of the Electrical Exhibition. M. Alphaud has given his adhesion to the request, which will be sent with his recommendation to the Commission of Sewers appointed by the Municipal Council, that when the Exhibition shall be closed, the railway will be kept running in the Champs Elysées.

About twenty minutes to eleven on Monday night, owing to some accident at present unexplained, the electric lights on the Brush system, one of the three with which experiments are at present being made in the City, were suddenly extinguished, leaving a large portion of the City in total darkness. The area over which the Brush light has been placed extends from Blackfriars Bridge, up Ludgate Hill, to St. Paul's Churchyard, and down Cheapside as far as Queen Street and King Street. Fortunately the old gas-lamps remain in their places while the electric light experiments are being made, and orders were quickly given for these to be lighted. Every attempt was made by those in charge of the Brush light to restore the connection, and for a few minutes it seemed as though they had succeeded; but this only lasted for a very short time, and it was soon seen that something had gone hopelessly wrong.

from *Nature* 24, 5 May, 11, 17, 1881

Carbon in carbonaceous chondrites: carbyne or graphite?

from P. P. K. Smith

CARBON in carbonaceous chondrite meteorites has received much attention as a likely carrier of isotopically anomalous noble gases. Over the past year this has led to a difference of opinion concerning the structural form of the carbon in such meteorites. This problem was addressed by several contributors to the Twelfth Lunar and Planetary Science Conference, held in Houston, Texas, in March.

Carbonaceous material has been reported^{1,2} as a carrier of part of the anomalous neon component, Ne-E, which is characterised by its extreme enrichment in the isotope ²²Ne. In 1969 Black and Pepin³ argued that Ne-E is of presolar origin, having been trapped in interstellar dust grains that were subsequently incorporated into the solar nebula. A carbonaceous host phase also seems likely¹ for a presolar xenon component, s-process Xe, that was discovered by Srinivasan and Anders⁴. Thus if the carrier phases of these exotic noble gas components can be identified they may provide us with examples of interstellar dust grains.

Last year Whittaker and Watts of the Aerospace Corporation, in collaboration with Lewis and Anders of Chicago University, reported that 80 per cent or more of the carbon in the Allende carbonaceous chondrite is in the form of carbynes⁵. Carbynes are a relatively recently discovered series of polymorphs of elemental carbon, based on a chain structure with alternating single and triple bonds⁶. Whittaker *et al.* identified carbynes in Allende by means of electron diffraction, and also by the characteristic negative-ion mass spectrum that they produce in an ion probe. Carbynes were also detected in the Murchison carbonaceous chondrite, where they were found to be carriers of Ne-E and s-process Xe.

A different view of the structure of the Allende carbon has been presented by Smith and Buseck of Arizona State University, who used high-resolution transmission electron microscopy (HRTEM) to obtain images of the microstructure of the carbon^{7,8}. Such images display prominent crystal lattice fringes with spacings between 3.4 and 3.9 Å, corresponding to the interplanar spacings of poorly crystalline graphite. The individual graphite crystallites are only ~100 Å in size, so that electron diffraction patterns obtained from areas 0.5 µm across (the smallest area that can be isolated by conventional selected-area electron diffraction) consist of diffuse rings. This

may be contrasted with the sharp, single-crystal diffraction patterns that Whittaker *et al.* obtained from carbyne crystals which were up to 0.75 µm in size.

At the recent conference Smith and Buseck reported an attempt to resolve this discrepancy⁹. The Chicago/Aerospace group had provided a portion of their nearly pure carbon separate (BK) from Allende, together with an electron microscope mount that they had prepared from this material. Smith and Buseck located several grains on this electron microscope mount that gave carbyne-like diffraction patterns. However, chemical analysis of these grains by means of X-ray microanalysis demonstrated that they were all sheet silicates rather than elemental carbon. Examination of an electron microscope mount of BK that Smith and Buseck prepared themselves revealed only graphitic carbon. Thus Smith and Buseck inferred that some, and possibly all, of the diffraction patterns that Whittaker *et al.* obtained were actually due to contamination of their sample by sheet silicates. This would appear to account for one unexplained feature of the Chicago study: although BK was considered to be 80 per cent carbyne, only five carbyne crystals were found in it by electron diffraction.

Of course these observations do not have any direct bearing on the ion-probe results of Whittaker *et al.*, and in discussion of the Arizona study Anders emphasised the striking predominance of even-numbered carbon species (C₂, C₄, etc.) that they had found in the negative-ion spectrum of the Allende material. This pattern is believed to reflect the tendency of the . . . C-C≡C-C . . . carbyne chain to cleave at the single bonds rather than at the triple bonds¹⁰. This feature is not in fact unique to carbynes — a predominance of C_{2n} species has been described previously for mass spectra obtained from both graphitic¹¹ and amorphous¹² carbon — but Whittaker *et al.* used reference spectra from graphite and from known carbynes to determine the carbyne content of their Allende carbon separate. The term graphitic carbon embraces a wide variety of microstructural forms, ranging from macroscopic single crystals to the extremely fine-grained tangled assemblage of ribbon-shaped crystallites that constitutes 'glassy' carbon¹³. Further work is perhaps needed to determine whether these varieties give rise to differing negative-ion spectra.

Electron diffraction observations of Allende carbon reported by Lumpkin¹⁴ are consistent with the HRTEM observations of graphitic carbon. Lumpkin did find a small fraction (~3 per cent) of grains that gave carbyne-like diffraction patterns, but

the chemistry of these grains has not yet been verified by microanalysis.

Dziczkaniec *et al.* described an investigation of the trapping of noble gases during the synthesis of carbonaceous material in a plasma discharge, an experiment designed to model the conditions under which the carbonaceous carrier phases in meteorites may have formed¹⁵. Although at an early stage, this study has shown that isotopic fractionation takes place during the trapping of xenon. It will be interesting to see whether the release temperatures for noble gases in this synthetic carbon follow a pattern similar to that found for meteoritic carbons. Interestingly enough, Dziczkaniec *et al.* found some crystalline grains amongst their run products that were very tentatively identified as carbynes. If this identification is confirmed, then this study would provide a new synthesis technique for carbynes, which have usually been prepared by laser heating of graphite.

Most recent studies of carbonaceous material in meteorites have centred on carbon-rich residues obtained by dissolving the bulk of the meteorite in concentrated acids, thereby destroying all information on the petrographical distribution of the carbon. An exception to this was the discovery by Scott *et al.* of aggregates of graphite and magnetite (Fe₃O₄) occurring as millimetre-size inclusions in several ordinary chondrite meteorites, and also in some more primitive unequilibrated ordinary chondrites¹⁶. The presence of this graphite-magnetite mixture in such primitive meteorites, which are thought not to have undergone appreciable heating in their asteroid parent bodies, indicates that graphite and magnetite may have been significant constituents of the early solar nebula. Textural relationships in these meteorites suggest that much of the

- Lewis, R. S., Alaerts, L., Matsuda, J. & Anders, E. *Astrophys. J.* **234**, L165 (1979).
- Eberhardt, P., Jungck, M. H. A., Meier, F. O. & Niederer, F. *Astrophys. J.* **234**, L169 (1979).
- Black, D. C. & Pepin, R. O. *Earth planet. Sci. Lett.* **6**, 395 (1969).
- Srinivasan, B. & Anders, E. *Science* **201**, 51 (1978).
- Whittaker, A. G. *et al. Science* **209**, 1512 (1980).
- Whittaker, A. G. *Science* **200**, 763 (1978).
- Smith, P. P. K. & Buseck, P. R. *Meteoritics* **15**, 368 (1980).
- Smith, P. P. K. & Buseck, P. R. *Science* (in the press).
- Smith P. P. K. & Buseck, P. R. *Lunar planet. Sci.* **12**, 1017 (1981).
- Stuckey, W. K. & Whittaker, A. G. *10th Carbon Conference*, Bethlehem, Pa. (1971).
- Honig, R. E. *J. chem. Phys.* **22**, 126 (1954).
- Fürstenau, N., Hillenkamp, F. & Nitsche, R. *Int. J. Mass Spectrom Ion Phys.* **31**, 85 (1979).
- Jenkins, G. M., Kawamura, K. & Ban, L. I. *Proc. R. Soc. Lond. A* **327**, 501, (1972).
- Lumpkin, G. R. *Lunar planet. Sci.* **12**, 631 (1981).
- Dziczkaniec, M., Lumpkin, G. R., Donohoe, K. & Chang, S. *Lunar planet. Sci.* **12**, 246 (1981).
- Scott, E. R. D. *et al. Lunar planet. Sci.* **12**, 955 (1981).
- Niederer, F. & Eberhardt, P. *Meteoritics* **12**, 327 (1977).

P. P. K. Smith is a Postdoctoral Research Associate in the Department of Geology, Arizona State University.

metallic iron present in them formed by the reduction of magnetite by carbon.

A previous report¹⁷ of presolar Ne-E in a carbon-rich residue from one of the ordinary chondrites studied by Scott *et al.* led them to suspect that the graphite-magnetite inclusions might be the carrier, and hence be of presolar origin. Unfortunately the story is not that simple — they

did not find any Ne-E in the graphite-magnetite intergrowths. Thus the Ne-E must either be contained in a different carbonaceous component, or in another minor phase in the residue. The identification of the carriers of such minor gas components is fraught with difficulties, but is providing us with valuable clues to the early history of the Solar System. □

Impact cratering

from Richard A.F. Grieve

The relative importance of cratering in planetary evolution was evidenced at the Twelfth Lunar and Planetary Science Conference* by numerous papers on impact craters and discussion, in sessions dealing with particular terrestrial planets and the icy satellites of Jupiter and Saturn, of their effects. The diversity of research reported in the sessions dealing directly with impact processes reflected the random nature of cratering with respect to specific geological environments and the involvement of many sub-disciplines of the physical science in providing information on cratering processes. As has been apparent in recent years, there was an increasing combination of theoretical and observational results, with each approach serving to constrain the other.

With respect to the evolution of the jovian and saturnian satellites, the results of impact experiments showed that the mechanical properties of ice differ from those of rock, even at the low temperatures of 70–80K existing on these bodies. In particular, the specific energy required to disrupt an icy body by impact appears to be ~ 20 per cent less than for a silicate body. Other impact experiments illustrated that extensive fracturing occurs around craters in ice but craters in ice-saturated sand targets exhibit more limited fracturing and are more like those in competent rock. It was noted, however, that ejection patterns appear to be similar for craters in icy and silicate targets.

Computational simulations derived from the solution of continuum mechanics equations continued to provide important insights into cratering processes. For example, they indicated that excavation dynamics and the physical state and radial distribution of ejecta are strong functions of impact velocity. Calculations also supported the interpretation that bowl-shaped cavities are produced by excavation plus the displacement of the cavity floor by impact-induced flow fields. Further confirmation of these displacements was proved by an analysis of the Cactus nuclear event. Similar conclusions on the formation of bowl-shaped craters were reached in a

combined observational and theoretical treatment of the Brent Crater, Canada, which also provided an estimate of the initial impact conditions: a 113-m radius chondritic projectile impacting at $\approx 27 \text{ km s}^{-1}$.

The important question of the detailed particle flow fields created in the target rocks by impact was addressed in a comparison of flow fields derived from continuum mechanics calculations and the 'Maxwell-Z' model. It was clear that this empirically derived model remains a powerful analytical tool. Similarities between the calculated flow fields for Meteor Crater, Arizona, and those of the Middle Gust III explosive event also suggested that the parametric codes derived from the latter may be applicable to impact events. However, the energy coupling differences suggest that explosive events are imperfect analogues to impact events in terms of scaling crater dimensions with energy.

It is known that the initial transient cavities formed by impact events generally modify rapidly to more stable final forms. In bowl-shaped craters, this modification is thought to occur by failure of the transient cavity walls. It was proposed at the meeting, however, that cavity modification may result from the centripetal motion of the entire rim area, in a manner similar to that suggested for the formation of larger, relatively shallow complex impact structures. The formation of these complex structures was addressed in a phenomenological model, ascribing central peak and ring formation to oscillatory motions of the transient cavity floor. These oscillatory motions require essentially plastic behaviour and viscosity of 10^9 P for the target rocks. Such viscosities may result from the effective fluidization of fractured target by an impact-generated acoustic field. Some support for the importance of two parameters, rock strength and gravity, in the phenomenological model may be found in the report that the 20-km Haughton impact structure on Devon Island, Canada, has a multi-ring form. Haughton was formed in sediments and its relatively small dimensions for a ring structure scale

with gravity and rock strength when compared with similar crater forms on the Moon and in crystalline targets on Earth. Doubts were raised, however, over the universal concept of deep initial transient cavities by the results of impact experiments at high gravitational accelerations, which suggested that relatively shallow cavities may be produced directly without modification.

Several papers dealt with the mineralogical or shock metamorphic effects produced by impacts in various target media. Shock experiments in limestone indicated that incipient vapourization of carbonate occurs at pressures as low as 350 kbar. An analysis of effects in tectosilicates illustrated that planar deformation features are produced during shock compression. It was also reported that the reduction in refractive index of shocked quartz is due to lattice disordering. A high concentration of lattice defects, particularly non-bridging oxygens, is also the cause of the conversion of plagioclase to the solid-state glassy phase known as maskelynite at shock pressures of 300–350 kbar.

The non-isochemical alteration of impact melt rocks was reported, as was evidence for extensive chemical exchange between feldspar clasts and enclosing impact melt matrix. This chemical exchange may account for some of the disturbances seen in the Rb–Sr systematics of melt rocks. ^{39}Ar – ^{40}Ar systematics of shocked samples from the Ries, Germany, were also described and it appears that up to 99 per cent of radiogenic ^{40}Ar may be lost at shock pressures of 300 kbar or more. This observation should aid in the interpretation of Ar ages obtained on lunar impact breccias.

A report of siderophile abundances in the melt rocks at Strangways suggested that this 20–25 - km impact structure in Australia was formed by an olivine-rich achondrite. Siderophile enrichments were also discussed with reference to the Cretaceous–Tertiary boundary layer and extinctions. It was noted that several important geochemical problems with respect to the various sampled sections need to be resolved. Theoretical calculations were presented for the postulated extinction-causing impact and indicated that significant atmospheric heating could only be achieved indirectly through energy transfer from ejecta. The interpretation that the formation of the 20-km lunar crater Giordano Bruno is recorded in the Gervase of Canterbury chronicles as occurring on 19 June 1178 was reviewed. It appears that only an estimated 0.1 per cent of the ejecta might reach the Earth by direct trajectories and that the analytical problems associated with recognizing such material are formidable. □

Richard A.F. Grieve is Visiting Professor at the Department of Geological Sciences, Brown University, Rhode Island.

*The Twelfth Lunar and Planetary Science Conference was held in Houston, Texas in March.

ARTICLES

The structure of West Antarctica from geophysical studies

E. J. Jankowski & D. J. Drewry

Scott Polar Research Institute, University of Cambridge, Lensfield Road, Cambridge CB2 1ER, UK

The subglacial topographical and geological configurations of West Antarctica and the boundary zone between East and West Antarctica are poorly known, yet they are crucial to our understanding of the tectonic evolution of the Pacific–proto Atlantic margin of Gondwanaland. Preliminary results are presented of simultaneous airborne radio echo and magnetic sounding of $\sim 10^6$ km² of West Antarctica including detailed maps of ice sheet and bedrock surfaces. Magnetics, radio echo and other available geophysical data suggest interpretations of sub-glacial geology.

DURING the late 1950s and 1960s major US oversnow traverses in West Antarctica conducted altimetric, seismic, gravimetric and magnetic studies^{1–4}. This work provided the first outline of ice sheet surface, subglacial topography and ice thickness. Many features were identified, including the principal ice divides, subglacial basins and mountainous blocks. Airborne magnetometry was also conducted, principally by the University of Wisconsin^{5,6} the US Geological Survey⁷ and US Project Magnet^{8,9}.

The oversnow traverses accomplished some 2,000 surface altitude determinations, ~ 250 seismic soundings and $\sim 2,000$ gravity observations in an area of $\sim 10^6$ km². Some interpolation and extrapolation was necessary for maps to be produced with a contour interval^{10,11} of 500 m. Although the aeromagnetic data had high accuracy, they suffered from having no simultaneous ice thickness information and only dead reckoning navigation.

During the austral summers 1977–78 and 1978–79 the Scott Polar Research Institute (SPRI) undertook an airborne geophysical programme in West Antarctica which included radio

echo sounding (by SPRI and Technical University of Denmark, TUD) and magnetometry (by Applied Physics Laboratory, APL, Johns' Hopkins University). The work was carried out in cooperation with the US National Science Foundation (NSF). In addition radio echo sounding (RES) alone was conducted during the austral summer 1974–75 in the area between Byrd Station and the Ross Ice Shelf¹² (see Fig. 1 for locations). The principal objective was to establish a grid for geophysical investigation of the complex structure of West Antarctica and its relationship to the East Antarctic craton. The available geophysical data combined with reconnaissance geology on scattered outcrops suggest West Antarctica possesses a complicated tectonic history^{13,14}.

Equipment during both seasons comprised pulsed radar systems operating at centre frequencies of 60 and 300 MHz, designed and built by TUD in collaboration with SPRI^{15,16}. Returned echoes were displayed in both A- and Z-modes and recorded on 35 mm film and dry-silver paper by a fibre-optic oscillograph. Equipment was mounted on pallets in a specially configured NSF C-130R Hercules aircraft, operated by US Navy Antarctic Development Squadron Six. An LTN-51 inertial system provided navigation data updated by photographic resections on known topographic points. Navigation and all other flight data were recorded on magnetic tape by an Airborne Research Data System (ARDS) constructed by US Naval Weapons Center, and operated since 1978 by APL. A proton precession magnetometer mounted in a tail stinger was installed by APL and operated by them simultaneously with RES. Coefficients compensating for the effect of the aircraft were determined by APL using a vector magnetometer and adopting the technique originally developed by Leliak^{17–19}. The unreduced magnetic data, provided by the NSF, were then compensated, reduced and correlated with RES in Cambridge.

Aircraft pressure altitude was corrected to points of known surface elevation. The terrain clearance was determined from RES records and the aircraft's radar altimeter. Navigation data were corrected for closure error, typically of the order of 3 km, and inflight fixes. Subsequent errors in altimetry from one flight to another were removed using a random walk technique. The mean error at crossover points before random walking was 46.5 m with a standard deviation of ± 44 m. RES records were digitized at a spacing of between 1.5 and 1.8 km and concatenated with updated navigation and altimetry data. Maps were compiled to depict ice sheet surface and bedrock configuration. In the case of the ice-sheet surface some 70,000 RES data points were used. Additional elevations were obtained by using all data from oversnow traverses (refs 1–4, 20 and C. R. Bentley, personal communication). For the bedrock map we used $\sim 56,000$ RES determinations combined with bedrock elevations at some 250 seismic stations on the traverses. A contour interval of 100 m has been used for the ice surface and 250 m for

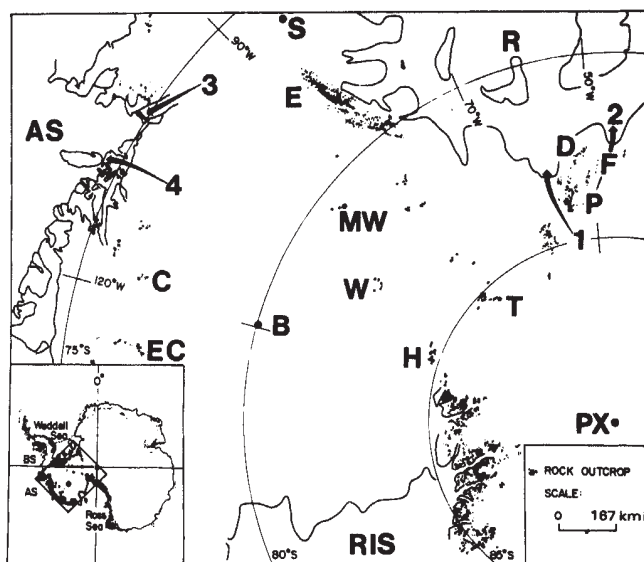


Fig. 1 Location map. AS, Amundsen Sea; BS, Bellingshausen Sea; R, Ronne-Filchner Ice Shelf; RIS, Ross Ice Shelf; B, Byrd Station; C, Crary Mountains; D, Dufek Massif; E, Ellsworth Mountains; EC, Executive Committee Range; F, Forrestal Range; H, Horlick Mountains; MW, Mount Woollard; P, Pensacola Mountains; PX, Pole Station; S, Siple Station; T, Thiel Mountains; W, Whitmore Mountains; 1, Foundation Ice Stream; 2, Support Force Glacier; 3, Pine Island Glacier; 4, Thwaites Glacier.

bedrock. Shading has been applied to areas of bedrock below -1,000 m a.s.l. to assist with interpretation. No attempt was made to detail the contours in areas of outcrop.

Ice sheet surface

Although there is broad correspondence between Fig. 2 and previous maps based on oversnow traverse^{20,21} or balloon data (tropical wind, energy conversion and reference level experiment, TWERLE)²², the large amount of data now available from RES allows considerably better delineation of many features including the ice drainage basins of West Antarctica. A major ice divide runs from inland of the Transantarctic Mountains through a gap between the Horlick and Thiel Mountains towards the Ellsworth Mountains confirming previous studies. A further ice divide crosses from the Executive Committee Range (passing to the north-east of Byrd Station) to join the first divide in the vicinity of Mt Woollard (see Fig. 1 for locations). West Antarctica is thus effectively divided into a series of drainage basins feeding ice into the Ross Ice Shelf, Ronne-Filchner Ice Shelf and the Amundsen-Bellinghousen Seas. Drainage into the Ross Ice Shelf from West Antarctica has been fully described elsewhere²³.

Ice streams shown shaded in Fig. 1 have been delineated by the presence of surface crevassing detected on the RES

records²³. The Foundation Ice Stream, running along the west side of the Pensacola Mountains, is already known in general but additional detail shows it to be a much more significant feature of West Antarctic drainage. Another ice stream is located 250 km further to the north-west. Although a less pronounced feature with boundaries lacking strong definition, its crevassed, linear margins are, nevertheless, clearly visible on RES records and from the air. It extends inland only for a few hundred kilometres. The area between these two ice streams forms a peninsula and remains one of the least known margins of the Ronne-Filchner Ice Shelf. Elsewhere LANDSAT satellite imagery has allowed the grounding line and ice rises to be delimited with considerable accuracy^{24,25}. RES records in this area reveal an undisputed ice-water interface indicating ice shelf. This is supported by ice thickness-surface elevation relationships. Further inland characteristic ice-rock reflections showing rapid fluctuations in instantaneous returned power are observed. Between the ice streams, however, a curious 'intermediate' zone has also been detected where received power is high and the echo return is smooth over considerable distances (tens of kilometres), yet ice thicknesses and surface elevations do not demonstrate hydrostatic equilibrium. Furthermore, high-resolution photographs from the Soviet 'Meteor' weather satellite show no clear grounding line shadow as exhibited elsewhere

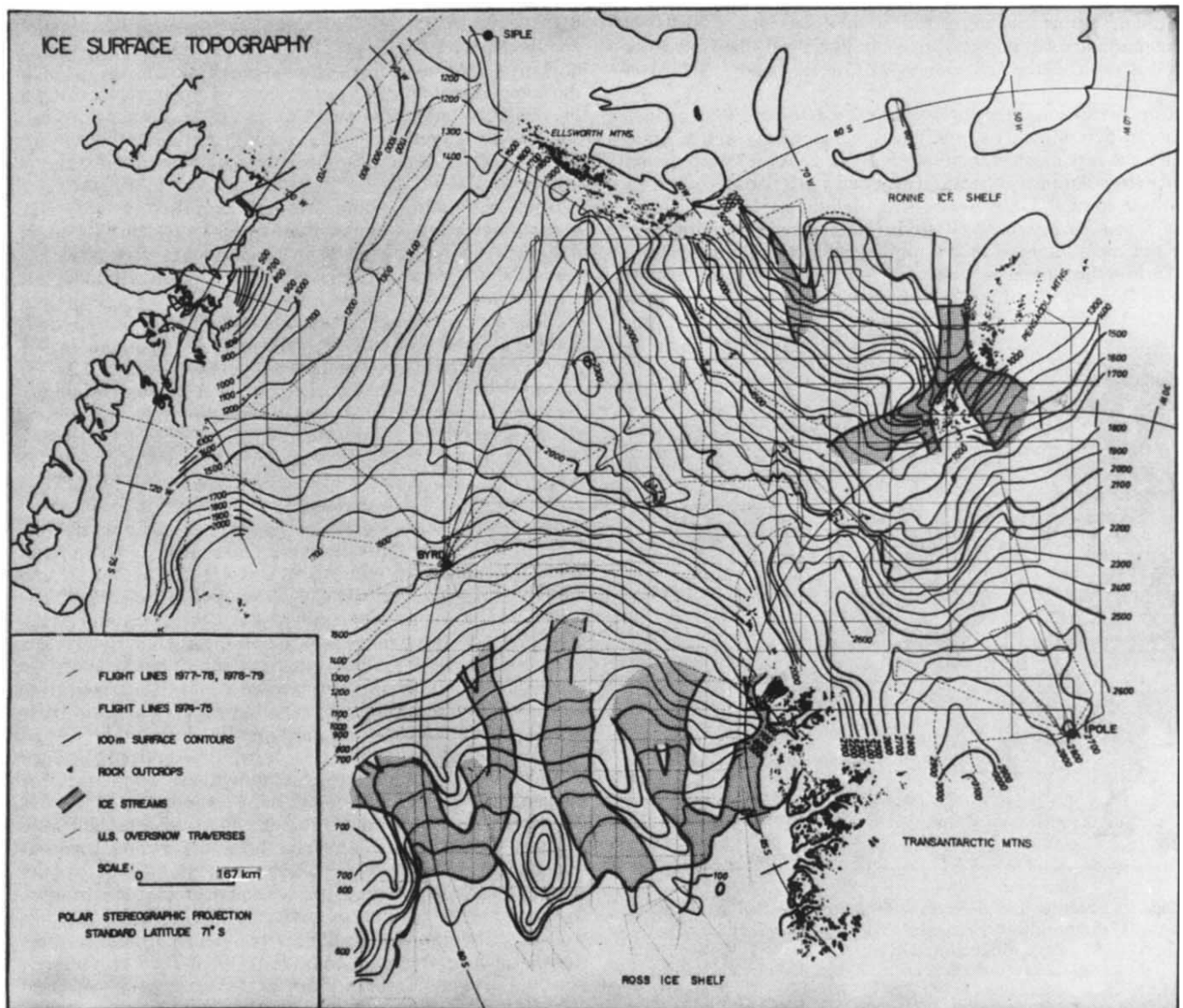


Fig. 2 Ice surface topography.

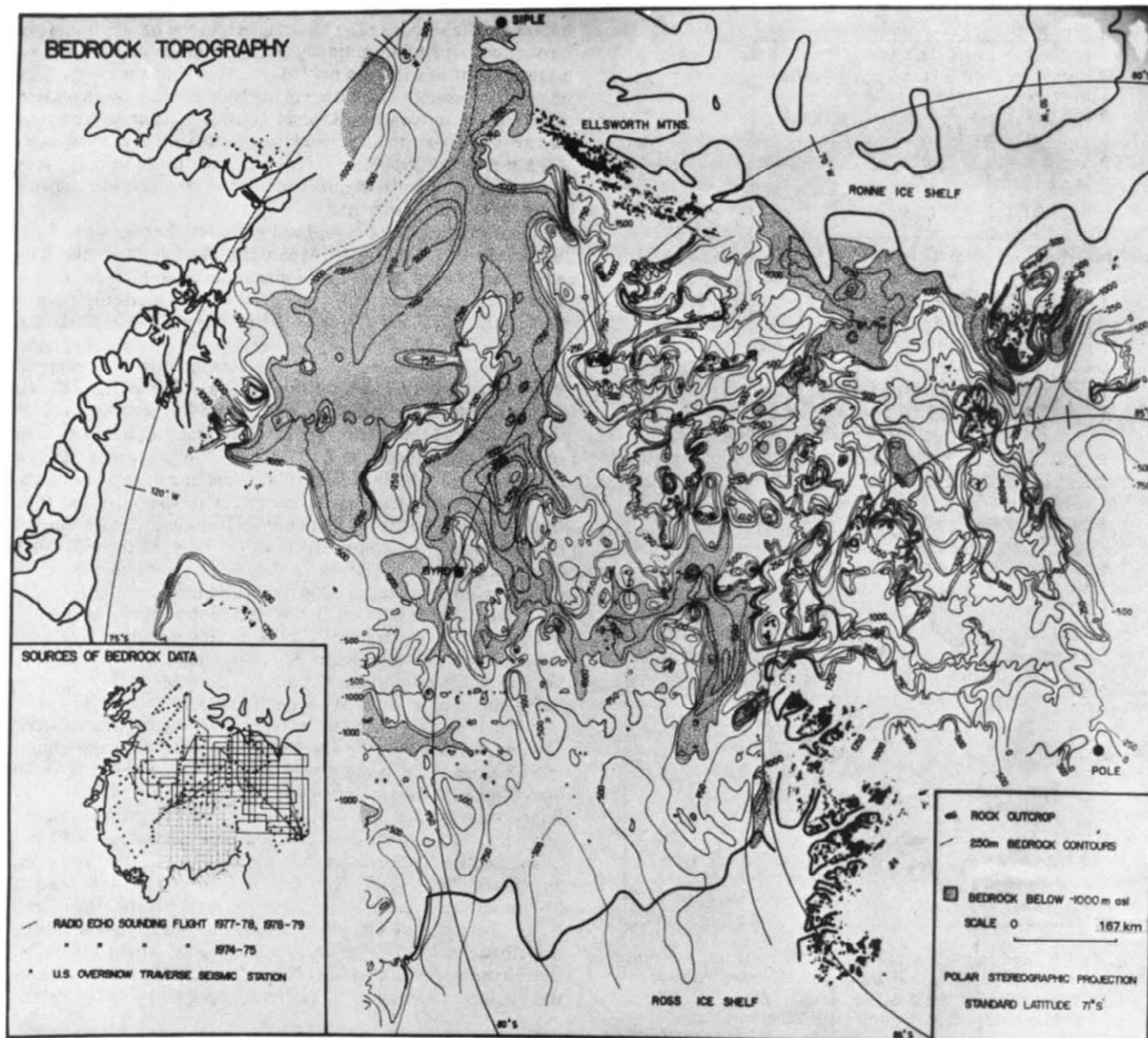


Fig. 3 Bedrock topography.

on the Ronne-Filchner Ice Shelf. The lack of relief-reflectivity further suggests that the area may possess an unusual and diffuse boundary²⁶. We explain the 'intermediate' zone as ice shelf resting on sea floor composed of soft water-saturated sediments—a similar phenomenon to that observed by Soviet glaciologists on the Lazarev ice shelf in 1975²⁷.

Ice drains into the Amundsen Sea principally through the Thwaites and Pine Island glaciers. Unlike the ice streams flowing into the Ross and Ronne-Filchner Ice shelves these major outlets are not protected by large ice shelves and unless they flow over high bedrock sills may be in a finely balanced state of equilibrium²⁸ or may not even be in equilibrium at all³⁷. RES suggests that Thwaites glacier does not have a high bedrock sill. There is, however, no pronounced erosional channel associated with the glacier which suggests that it may be a transient phenomenon. No data were collected on the Pine Island Glacier.

Bedrock configuration

The map presented in Fig. 3 is the first detailed and systematic study of bedrock using RES in West Antarctica. There is again a broad correspondence between major topographical features and previous maps based on seismic soundings and interpretation of free air gravity anomalies^{10,11}. The location of all seismic stations used in this survey are shown in Fig. 3.

Compared with previous work, however, the most striking feature of the map (and profiles) is the considerable improvement in topographical detail.

The area mapped can be divided into several parts some of which had been known from earlier work, and are briefly outlined here. (1) At the head of the Ross Ice Shelf the terrain is very smooth with subdued ridges and troughs marking the position of the ice streams draining into the ice shelf. This area is terminated to the north-east by (2) an extensive subglacial basin (Byrd Subglacial Basin). Insufficient data may account for its apparent smoothness. (3) A completely new feature is the sinuous ridge running across the centre of the basin. Near Byrd Station it is prominent on RES records, standing 1,000 m above the floor of the basin. (4) At the northern margin of the basin Cenozoic volcanic rocks crop out in the Cray Mountains and Executive Committee Range. RES records show these areas of outcrop to be steep sided and separated by deep narrow troughs. (5) A major highland region extends southwards from the exposed Ellsworth Mountains. The terrain is extremely rugged, possessing a relief amplitude of 2 km, a feature not previously apparent nor emphasized on the map due to spacing of flight lines and the contour interval. There is a very well defined boundary between this highland area and the Byrd Subglacial Basin. The Whitmore Mountains form a steep-sided upstanding

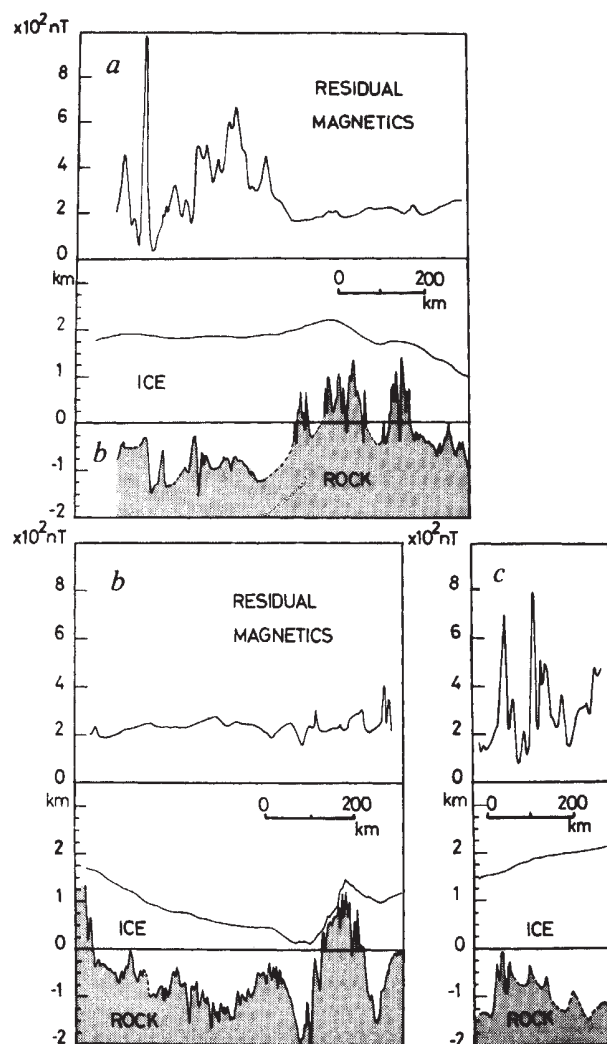


Fig. 4 RES and magnetic profiles in West Antarctica. *a*, Boundary between Byrd Subglacial Basin and Ellsworth Mountains block. *b*, Bedrock low at head of Ronne-Filchner Ice Shelf. *c*, Sinuous ridge cutting across Byrd Subglacial Basin.

massif separated from the Ellsworth Mountains block by a narrow trough. Other scattered outcrops to the south of the Ellsworth Mountains appear to be part of the same mountainous region. (6) The Transantarctic Mountains constitute a separate highland area and form a semi-continuous chain across Antarctica. Two deep valleys almost sever the Horlick Mountains from the rest of the Transantarctic Mountains, whereas the main part of the Pensacola Mountains is bounded on two sides by very deep narrow troughs occupied by Foundation Ice Stream and Support Force Glacier. At the northern end of the Pensacola Mountains there is a small depression between the exposed Forrestal Range and Dufek Massif. (7) At the head of the Ronne-Filchner Ice Shelf bedrock forms a deep depression which widens towards the ice shelf.

Geological interpretation

Bedrock at the head of the Ross Ice Shelf has been fully discussed elsewhere¹² where subdued topography and magnetic signatures are attributed to a substantial thickness of late Mesozoic-Cenozoic sediments extending the Ross Sea sedimentary basin beneath the grounded ice sheet. Further inland there is a significant change in magnetic character towards the volcanic outcrops in Northern Byrd Land and the Byrd Subglacial Basin, as noted in previous surveys^{1,4-6}. Large anomalies ($\sim 1,000$ nT) are here associated with outcrops and subglacial

features. The sinuous ridge running across the basin possesses a pronounced magnetic anomaly (Fig. 4c): modelling in progress suggests that this feature may also have a volcanic origin. The magnetic character of the floor of the Byrd Basin shows a pattern of irregular anomalies. Seismic refraction studies from the centre of the basin indicate ~ 2 km of strata with P-wave velocity 4.1 km s^{-1} (ref. 29). These results suggest that the basin may constitute a sequence of interbedded sedimentary and volcanic strata overlying basement.

The pronounced topographical boundary between the Byrd Basin and rugged Ellsworth Mountains block is also marked by an abrupt change in pattern of magnetics from high-amplitude anomalies associated with the Basin to a magnetically quieter field with only isolated anomalies of wavelength ~ 40 – 50 km (Fig. 4a). The subdued magnetics are consistent with subglacial extension of a geological sequence similar to the exposed Ellsworth Mountains (Palaeozoic strata). This margin of the Ellsworth block becomes less well defined further west towards the Whitmore Mountains (south of Byrd Station); it is likely that sedimentation since the late Mesozoic has masked deeper structures and inhibits accurate identification of this boundary. A series of subglacial basins continues to the west and then trends south towards the front of the Transantarctic Mountains and may mark the continuation of the edge of the Ellsworth block. The Ellsworth block as described by the above boundaries includes an area around the Whitmore Mountains. These are thought to be geologically related to the Ellsworth chain³⁰, but may form a separate unit if the narrow trough between the two is structurally significant. Similarly, there is no clear break between the Whitmore Mountains and Thiel Mountains, the two being linked by a series of upstanding blocks.

Further south, magnetic profiles over the Transantarctic Mountains (except at the northern end of the Pensacola Mountains) suggest the presence of a significantly thick layer of non-magnetic rocks—possibly corresponding to Beacon Supergroup strata and older Precambrian metasediments. Modelling of anomalies just north of Thiel Mountains indicates a substantial sedimentary section²⁶. The largest recorded anomalies ($>1,000$ nT) are seen in the northern Pensacola Mountains (Forrestal Range and Dufek Massif). These two ranges form part of a large mafic intrusion which has been fully described elsewhere^{7,31,32}. The lopolith-like synclinal form of the intrusion, developed as a result of weak folding, accounts for the shallow depression in bedrock topography between the ranges⁷.

The distinctive bedrock low at the head of the Ronne-Filchner Ice Shelf lies between the Transantarctic and Ellsworth Mountains, widening into the Weddell Sea. Hochstein calculated depth to MOHO at $83^\circ \text{S } 70^\circ \text{W}$ on the eastern edge of this basin, reporting thin crust of 24 km thickness³³. Previous work has suggested that the Ellsworth Mountains and any associated subglacial block are out of place in West Antarctica based on their stratigraphy and structural trends^{14,34-36}. The thin crust lying beneath the bedrock low may thus represent crust rifted during the displacement of the Ellsworth block. Magnetic records from this basin area show no significant anomalies and confirm earlier suggestions of a substantial section of sedimentary rock in the area⁷ (Fig. 4b).

The present studies are being extended to provide quantitative estimates of rock magnetizations, depth of sedimentary cover, a detailed analysis of the structure and areal extent of the Dufek Massif (in collaboration with US Geological Survey), and to reconsider the position of West Antarctica in Gondwanaland.

This work was supported by NERC grant GR3/2291 and the NSF Division of Polar Programs. C. R. Bentley provided unpublished data from the Wisconsin traverse. We thank the NSF for aeromagnetic data, Dr J. C. Behrendt and the US Geological Survey Branch of Regional Geophysics for assistance in their reduction, and Drs G. de Q. Robin and D. H. Elliot for helpful discussions. TWERLE data were made available to SPRI by kind permission of N. Levanon and were plotted and contoured at SPRI for comparison with RES.

1. Bentley, C. R. & Ostenso, N. A. *Geophys. J.* **18**, 882-912 (1964).
2. Behrendt, J. C. *Geophys. Polar Res. Centre Rep.* 64-1 (University of Wisconsin, 1964).
3. Robinson, E. S. thesis, Univ. Wisconsin (1964).
4. Bentley, C. R. & Chang, F.-K. *Ant. Res. Ser.* **16**, 1-38 (1971).
5. Behrendt, J. C. & Wold, R. J. *J. geophys. Res.* **68**, 1145-1153 (1963).
6. Behrendt, J. C. *Science* **144**, 993-994 (1964).
7. Behrendt, J. C., Henderson, J. R., Meister, L. & Rambo, W. L. *US Geol. Surv. Prof. Pap.* 884 (1974).
8. *US Naval Oceanogr. Office, Spec. Publ.* 66 (1963).
9. *US Naval Oceanogr. Office, Spec. Publ.* 66, Suppl. 1 (1965).
10. Heezen, B. C., Tharp, M. & Bentley, C. R. *Am. geogr. Soc. N.Y.*, Folio 16 (Antarctic Map Folio Series, 1971).
11. Korotkevich, Ye. S., Koblents, Ya. P. & Kosenko, N. G. *Karta Korennogo Rel'yefa Antarktidy, 1:10000000* (Soyuzmorniprojekt, Arctic and Antarctic Institute, Leningrad, 1977).
12. Rose, K. E. thesis, Univ. Cambridge (1978).
13. Craddock, C. in *Antarctic Geoscience* (ed. Craddock, C.) (in the press).
14. Elliot, D. H. *Am. J. Sci.* **274A**, 45-106 (1975).
15. Skou, N. & Sondergaard, F. *TUD, Lab. Electromagn. Theor. Rep.* 169 (1976).
16. Christensen, E. *TUD, Lab. Electromagn. Theor. Rep.* 93 (1970).
17. Leliak, P. *IRE Trans. Aeron. Nav. Electron.* September (1961).
18. Finkel, A. *APL/S2R-78-194* (1978).

19. Hickerson, R. L. *APL/S2R-79-145* (1979).
20. *Am. geogr. Soc. N.Y.* 1:5000000, *Sheet 13, Antarctica* (1970).
21. *Ministerstvo Morskogo Flota CCCP Karta Antarktika* (1971).
22. Levanon, N., Julian, P. R. & Suomi, V. E. *Nature* **268**, 514–516 (1977).
23. Rose, K. E. *J. Glaciol.* **24**, 63–75 (1979).
24. Swithinbank, C. W. M., Doake, C. S. M. & Crabtree, R. D. *Polar Rec.* **18**, 295–299 (1976).
25. Crabtree, R. D. & Doake, C. S. M. *Polar Rec.* **20**, 31–37 (1980).
26. Drewry, D. J., Meldrum, D. T. & Jankowski, E. *Polar Rec.* **20**, 45–47 (1980).
27. Korotkevich, Ye. S., Savtyugin, L. M. & Morev, V. A. *Inf. Byull. sov. Antarkt. Eksped.* No. 98, 45–52 (1978).
28. Thomas, R. H., Sanderson, T. J. O. & Rose, K. E. *Nature* **277**, 355–358 (1979).
29. Bentley, C. R. & Clough J. W. in *Antarctic Geology and Geophysics* (ed. Adie, R. J.) 683–691 (Oslo Universitetsforlaget, 1972).
30. Craddock, C. & Webers, G. F. in *Antarctic Geoscience* (ed. Craddock, C.) (in the press).
31. Ford, A. B. *US geol. Surv. Bull.* 1405-D (1976).
32. Behrendt, J. C., Drewry, D. J., Jankowski, E. & Grim, M. S. *Science* **209**, 1014–1017 (1980).
33. Bentley, C. R. *Tectonophysics* **20**, 229–240 (1973).
34. Schopf, J. M. *Science* **164**, 63–66 (1969).
35. Stump, E. in *Implications of Continental Drift to the Earth Sciences* Vol. 2 (eds Tarling, D. H. & Runcorn, S. K.) 909–924 (Academic, London, 1973).
36. Clarkson, P. D. & Brook, M. *Nature* **265**, 615–616 (1977).
37. Stuiver, M., Denton, G. H., Hughes, T. & Fastook, J. L. in *The Last Great Ice Sheets* (eds Denton, G. H. & Hughes, T.) 319–435 (Wiley, New York, 1981).

C. E. Ford

Petrogenetic relationships between compositions advocated as possible liquids parental to the Skaergaard layered series cumulates have been investigated using equations for the prediction of liquidus temperature and liquid path followed during fractional crystallization. Given a suitably defined sequence of primocryst mineral assemblages, a variation of the fractionation scheme provides a new method for the calculation of the parent liquid composition.

The composition of the Skaergaard parent liquid is uncertain. Samples of the chilled marginal gabbro are heterogeneous² and may have been affected by wall rock contamination. Chill

Fig. 1 Projection from M_2S into part of the plane CS-MS-A of EG4507, KT39 and derived liquid compositions. At high pressure, crystallization of ol+pl+cpx from EG4507 drives the liquid composition away from the ol-pl-cpx tie-plane. At low pressure, crystallization of ol+pl precedes crystallization of ol+pl+cpx. Curve *b* represents negative fractional crystallization of ol+pl from KT39. This curve does not intersect those representing fractional crystallization of EG4507 at either low or high pressure—at the apparent points of intersection the mineral compositions are significantly different. Curve *c* is parallel to the liquid path followed in many cases of equilibrium crystallization of ol+pl+cpx.

Table 1 Composition of Skaergaard chills and three possible cumulus enriched bulk compositions

	KT39	EG4507	B1	B2	B3
SiO ₂	50.10	48.08	44.10	48.01	46.85
TiO ₂	2.64	1.17	0.59	0.59	0.59
Al ₂ O ₃	13.40	17.22	8.61	2.33	20.40
Fe ₂ O ₃	1.77	1.32	0.66	0.66	0.66
FeO	11.81	8.44	11.12	4.22	6.26
MnO	0.20	0.16	0.08	0.08	0.08
MgO	6.54	8.62	27.34	4.31	11.10
CaO	10.09	11.38	5.69	13.89	11.48
Na ₂ O	2.42	2.37	1.19	2.28	1.96
K ₂ O	0.57	0.25	0.13	0.13	0.13
P ₂ O ₅	0.22	0.10	0.05	0.05	0.05
H ₂ O ⁺	—	1.01	0.51	0.51	0.51
Total	99.76	100.12	100.07	100.06	100.07
Predicted liquidus mineral compositions			Predicted solidus mineral compositions		
O1	Fo 76(74)* Fo 86		Fo 82	Fo 65	Fo 76
P1	An 73(67)* An 81		An 62	An 74	An 73

KT39 from ref. 6, EG4507 from ref. 1. B1 = EG4507 + 50% olivine (Fo 86). B2 = EG4507 + 50% plagioclase (An 81). B3 = EG4507 + 15% olivine (Fo 86) + 35% plagioclase (An 81).

* Composition stated to be present in experimental charge.

cation% of fractional crystallization of chill EG4507 would be necessary to produce olivines of the composition observed at the exposed base of the layered series. However, geophysical evidence precludes the existence of such an extensive hidden zone immediately beneath the exposed layered series⁵. A parent liquid of the composition of the much more evolved, but less ubiquitous chill KT39 could perhaps produce the layered series with a minimal hidden zone⁶. The relationships between the various Skaergaard chill compositions and the layered series are thus not easily explained.

Evolving Skaergaard liquids and coexisting primocrysts

To calculate the extent of fractional crystallization required for a particular parent liquid to proceed to that which will precipitate a particular cumulus mineral assemblage the identity of the primary compositions of the cumulus minerals (primocrysts) are needed as the latter may have changed in composition during equilibrium crystallization of trapped liquid.

The Skaergaard layered series cumulus plagioclase crystals consist of homogeneous cores and zoned rims. The cores have been interpreted as primocryst compositions with the rims representing the crystallization product of trapped inter-cumulus liquid¹⁷. However, olivine crystals are virtually homogeneous⁸ and must have continuously re-equilibrated, and become less magnesian, during further crystallization of trapped liquid. The extent of the change in primocryst mineral compositions during re-equilibration depends on the proportions of primocrysts and trapped liquid. This may be illustrated by considering chill EG4507 which has olivine(Fo86) and plagioclase(An81) as liquidus phases⁹. In Table 1 compositions B1, B2 and B3 are the result of accumulation of 50% of olivine, 50% of plagioclase and 50% of olivine(15%) + plagioclase(35%) into the original liquid. The ratio of olivine to plagioclase incorporated into EG4507 to form composition B3 is that in which they are predicted to crystallize during fractional crystallization. For these bulk compositions the equilibrium mineral compositions at the solidus can be approximated by the CIPW norm with the anorthite and albite contents recast to indicate plagioclase composition and the Fe–Mg distribution between the ferromagnesian minerals reset using measured cation exchange reaction coefficients^{10,11}. Note that the olivine in the bulk composition buffered by the 50% of previously crystallized cumulus olivine is significantly more magnesian than the olivine in the composition buffered by 50% of cumulus plagioclase.

Similarly, plagioclase is much more calcic in the bulk composition equilibrated with 50% of cumulus plagioclase. Such variations in mineral composition do occur in layers as narrow as 10 cm (refs 8, 12).

A critical point in the fractional crystallization sequence of the Skaergaard parent liquid is that at which clinopyroxene first appears as a cumulus phase at the base of LZb. Identification of the primocryst mineral compositions at this point could constrain the composition of the coexisting liquid and thus of the range of liquids in the fractionation sequence. From this discussion, and the fact that plagioclases are zoned, the best assumption is that the coexisting primocryst compositions are the most calcic plagioclase and most magnesian ferromagnesian minerals observed at this structural height, not necessarily from the same sample: olivine(Fo64), plagioclase(An63) and clinopyroxene(Di71 = Mg/(Mg + Fe²⁺))^{1,6}. The problem is then to identify which of the possible parent liquids would precipitate this mineral assemblage at the point where clinopyroxene joins olivine and plagioclase as fractionating phases.

Theoretical fractional crystallization

To predict magmatic evolution based on known phase equilibria we need to predict liquidus temperatures, liquidus phase compositions and the change in liquid composition produced by perfect fractional crystallization of the liquidus phases. Nathan and Van Kirk¹¹ presented a generalized method incorporating these features but the accuracy of liquidus temperature equations, relating mineral phase liquidus temperature to liquid composition at atmospheric pressure, have been criticized¹⁷. Applied to a data set consisting of microprobe analyses of glasses produced in experiments at atmospheric pressure in this laboratory, and for which experiment temperature and oxygen fugacity were known, the equations produce systematic discrepancies of up to 40 °C. The data set has been compiled from experiments on 24 terrestrial basaltic rocks, 5 komatiites and 10 achondritic meteorite and lunar basalts. At present, the data set contains 336 olivine saturated glasses, 296 spinel–magnetite saturated glasses, 160 plagioclase saturated glasses, 116 clinopyroxene saturated glasses and 67 low Ca–pyroxene saturated glasses. With the equation compositional coefficients revised based on this data set, liquidus temperatures for olivine, plagioclase, clinopyroxene and magnetite fit the equations with a standard deviation of error of <6 °C (refs 18, 19). With this degree of accuracy in liquidus temperature prediction it is possible to be much more accurate in the calculation of fractionation paths and more specific in the interpretation of petrogenetic relationships.

The revised coefficients for the orthopyroxene liquidus temperature equation, derived from the orthopyroxene–glass pairs in the data set, give a standard deviation of error of <5 °C (all structural modifications of low-Ca pyroxene are referred to as orthopyroxene). However, the resulting equation predicts that orthopyroxene would be the liquidus phase in glass compositions which did not coexist with orthopyroxene. This anomaly probably results from the limited range of liquid compositions in the data set for orthopyroxene–glass pairs and illustrates the possible consequences of undue extrapolation of the liquidus temperature equations. During fractional crystallization in a multi-component system the extent of extrapolation may not be readily apparent. Until this anomaly is resolved the quantitative application of the mineral temperature equations will be restricted to equilibria not involving orthopyroxene.

The predicted and experimentally determined liquidus temperatures for EG4507 and KT39, neither of which were included in the data set, are presented in Table 2. For KT39 the equations predict that orthopyroxene would be close to the liquidus although in the experiments this phase was not detected after 55 °C of equilibrium crystallization⁶. Apart from this, the rather small discrepancies between predicted and experimentally determined liquidus temperatures may be accounted for by known experimental and analytical problems. The liquidus temperature equations are, therefore, considered sufficiently

accurate to demonstrate possible petrogenetic relationships between Skaergaard chill compositions before the appearance of orthopyroxene. It could even be argued that the predictions for KT39 fit the observed petrography⁶ better than the experiments.

In many pressure-temperature (P - T) phase diagrams of basic rocks clinopyroxene becomes the liquidus phase at elevated pressure. The possible consequences of the involvement of clinopyroxene in fractionation equilibria at high pressure can be considered after assessment of the effects of total pressure on the liquidus temperature equations. The effects of total pressure can be estimated approximately from the experimentally determined slopes of liquidus surfaces on P - T phase diagrams of rocks²¹⁻²⁵. Average values of $+4^\circ\text{C kbar}^{-1}$ for olivine, plagioclase and magnetite and $+9^\circ\text{C kbar}^{-1}$ for clinopyroxene and orthopyroxene have been adopted. Similarly the effects of small quantities of dissolved water can also be estimated from data in the systems albite-water and diopside-anorthite-water¹³. Values of $-250^\circ\text{C} \times H^+$ (cation fraction) for felsic minerals and $-80^\circ\text{C} \times H^+$ for mafic minerals have been adopted.

Using the mineral temperature equations to identify the liquidus phase the liquid path followed during fractional crystallization can be approximated by repeated removal of a small amount of the current liquidus phase using equations for exact fractionation¹⁴. For olivine, clinopyroxene and orthopyroxene these are of the form

$$\text{Fe} = (\text{Fe}_0/\text{Mg}_0^K) \cdot \text{Mg}^K$$

where the element values are absolute cation quantities in the liquid before (Fe_0 , Mg_0) and after (Fe , Mg) each fractionation step and K is the appropriate cation exchange reaction

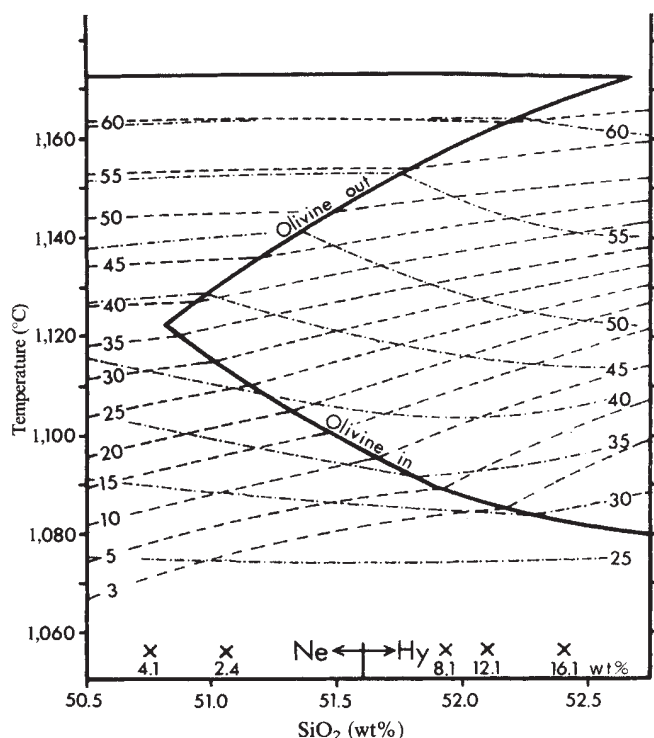


Fig. 2 Fractional crystallization diagram based on the liquidus-temperature equations from ref. 11 for the univariant range of liquids in the system SiO_2 - Al_2O_3 - FeO - MgO - CaO - Na_2O which all have olivine(Fo64), plagioclase(An63) and clinopyroxene(Di71) as simultaneous liquidus phases. Vertical lines illustrate the changing primocryst compositions and assemblages which would be produced by fractional crystallization. Dashed lines are isopleths of Fo content of olivine, shown with metastable continuation through the field in which olivine is absent, but orthopyroxene is present, while dashed-dotted lines are isopleths of An content of plagioclase. The nepheline and hypersthene scales are not linear.

coefficient^{10,11}. The latter may also be used to determine the composition of the liquidus phase. The amounts of other components to be removed from the liquid are determined by the stoichiometry of the phase in question. For plagioclase and magnetite the equations for exact fractionation can be solved by numerical approximation:

Plagioclase:

$$\text{Na}^{((K, \text{Ca}, \text{Al})/\text{Si})} = (\text{Ca} \cdot \text{Na}_0^{((K, \text{Ca}, \text{Al})/\text{Si})})/\text{Ca}_0$$

Magnetite:

$$\text{Ti}^{((K, \text{Fe}^{3+})/(\text{Fe}^{2+}, \text{Ti}))} = (\text{Fe}^{3+} \cdot \text{Ti}_0^{((K, \text{Fe}^{3+})/(\text{Fe}^{2+}, \text{Ti}))})/\text{Fe}^{3+}_0$$

At a phase boundary, crystallization of one phase results in the liquid moving into the primary phase volume of one of the other phases present. This then becomes the liquidus phase for the next fractionation step. A cotectic is thus followed by a zig-zag approximation path which is almost independent of the size of the crystallization increment provided that the increments are sufficiently low. At a reaction boundary the phase in reaction fails to become a liquidus phase and ceases to fractionate. Specification of a negative increment allows a liquid fractionation path to be followed back to higher temperature, more primitive compositions. This process can be described as negative fractional crystallization to avoid conflict of terminology with melting processes.

The liquidus temperature and fractionation equations have been incorporated into a computer program. Given a parent liquid composition the program has been designed to list the compositions and proportions of crystallizing phases and the evolved liquid composition at regular intervals during the fractionation process. The compositions of crystallizing phases correspond to the primocryst mineral compositions of cumulate sequences.

Crystallization of Skaergaard chills

At atmospheric pressure, 37% of fractional crystallization of olivine and plagioclase from EG4507 is predicted to occur before the appearance of clinopyroxene. Only 3% of fractional crystallization of olivine from KT39 is predicted to occur before plagioclase and clinopyroxene appear almost simultaneously. In both cases the ferromagnesian minerals are more magnesian and plagioclases more calcic than the quoted primocryst compositions at the base of LZb. Neither composition could apparently be directly parental to the layered series by fractional crystallization at low pressure unless re-equilibration of primocrysts with trapped liquid has resulted in an incorrect identification of the primocryst compositions.

EG4507 and KT39 are shown in Fig. 1 as an M2S projection into the plane CS-MS-A (ref. 15). The liquid path followed during fractional crystallization of EG4507 at low pressure does not pass through, or even close to, the composition of KT39. Note specifically the orientation of the liquid path followed during fractional crystallization of olivine + plagioclase + clinopyroxene. This is significantly different from the path followed during equilibrium crystallization of these phases. Least squares mixing calculations using the method of Wright and Docherty¹⁶ demonstrate that KT39 could be combined with the olivine, plagioclase and clinopyroxene compositions present close to the liquidus to produce EG4507 (45.1% KT39 + 12.6% olivine(Fo74) + 35.6% plagioclase(An73) + 6.7% clinopyroxene(Di79) = EG4507 sum $r^2 = 0.5$). KT39 could thus be a liquid produced during equilibrium crystallization of a liquid such as EG4507 before separation of crystals from liquid. However, this is not an essential relationship.

Cumulate sequence parent liquids

Although it might be possible to find a chill composition able to produce the observed layered series primocryst assemblages during fractional crystallization it is possible to adopt a radically new approach to the calculation of parent liquid compositions.

Let an olivine-gabbro mineral assemblage consist of primocrysts of olivine (Fo64), plagioclase (An63) and clinopyroxene

Table 2 Comparisons of experimentally determined and calculated liquidus temperatures (°C) for Skaergaard chills at atmospheric pressure

Mineral	EG4507 (f_{O_2} = iron/wustite)*			KT39 (f_{O_2} = magnetite/wustite)*		
	XP†	F‡	N & V§	XP	F‡	N & V§
Olivine	1,230–1,248	1,225	1,241	1,171–1,179	1,179	1,187
Plagioclase	1,230–1,248	1,219	1,220	1,171–1,174	1,167	1,165
Magnetite	<1,174	1,193	—	1,171–1,179	1,154	1,142
Clinopyroxene	1,188–1,208	1,173	1,213	1,134–1,140	1,157	1,181
Orthopyroxene	<1,174	1,181	1,166	<1,115	1,176	1,170

The liquidus-temperature predictions for phases which would not actually be the liquidus phases cannot be compared with the temperature of appearance of these phases during equilibrium crystallization. They are given for comparison and to give an approximate indication of the possible effect of total pressure (see text).

* Fe_2O_3 and FeO contents recalculated using an equation given in ref. 11.

† From ref. 9. Upper temperature is that at which phase was absent, lower temperature is that at which phase was present.

‡ F, prediction based on data set of experimental glasses.

§ N & V, prediction of the equations given in ref. 11.

|| From ref. 6, magnetite liquidus may be too high.

(Di71), the assemblage marking the base of LZb. In the system $SiO_2-Al_2O_3-FeO-MgO-CaO-Na_2O$ one liquid which would precipitate this assemblage would lie in the pseudo-ternary plane olivine(ForX)-plagioclase(AnY)-clinopyroxene(DiZ), where X, Y and Z represent the mineral compositions which form the plane end-members. Values for X (=Z) can be calculated from crystal-liquid iron-magnesium cation exchange reaction coefficients^{10,11} and the identified primocryst ferromagnesian mineral compositions. The value of Y depends on the proportions of plane end-members contributing to the liquid composition at the piercing point where the three crystalline phases are in equilibrium with liquid. It is easy to estimate Y, use the liquidus temperature equations to generate the pseudo-ternary phase diagram, observe the composition of the plagioclase in equilibrium with the ferromagnesian minerals of the selected composition, and refine Y accordingly. The liquid composition is represented by the plane end-members at the piercing point determined to closer than 0.1 wt%.

The resulting liquid composition is not unique. In a six-component system the phase rule is uniquely satisfied by the coexistence of four phases and the specification of pressure and the compositions of three of the phases. However, in this instance the compositions of olivine and both pyroxenes are determined by the $Mg/(Mg + Fe)$ ratio of the liquid. Therefore, this mineral assemblage could be precipitated from a univariant range of liquid compositions in the six-component system. The extent of this range can be determined by altering the concentration of any one of the components of plagioclase in the parent liquid. For convenience SiO_2 has been chosen because the effect of varying the concentration of this component is easily expressed in terms of normative nepheline, olivine, hypersthene or quartz.

Figure 2 is a fractional crystallization diagram in which the specified mineral assemblage is crystallized at the liquidus over a range of nepheline and hypersthene normative compositions. Note that different liquids would produce progressively differing primocryst assemblages as illustrated by the mineral composition isopleths and the location of the boundaries defining the field in which olivine is absent. This is the key to the interpretation of natural igneous cumulate assemblages. The extent of the orthopyroxene field and its extension into nepheline normative compositions are almost certainly erroneous, resulting from extrapolation of the orthopyroxene liquidus temperature equation. Note also that the extent of the 'olivine gap' is a function of the SiO_2 content of the parent liquid.

Adding more components to the system will produce a greater range of liquids capable of precipitating the same mineral assemblage. However, it should be possible to restrict element abundances in parent liquids from consideration of crystal-liquid partition coefficients or from element abundances in extrusive rocks. The one liquid which is best able to reproduce the observed sequence can be obtained by adjustment of the minor components by trial and error. To promote the appearance of cumulus magnetite at the correct point in the

fractionation sequence in relation to the compositions of the coexisting primocrysts at the base of LZc there is comparatively little room for adjustment of the contents of K_2O , TiO_2 and Fe_2O_3 in the parent liquid. However, as shown in Fig. 2, it is the point of disappearance and reappearance of olivine, as a result of reaction relationships with orthopyroxene, and the compositions of primocrysts in the later stages of fractional crystallization which identify more clearly the most probable parent liquid composition. Refinement of the liquid composition present at the base of LZb should, therefore, await additional experimental data on orthopyroxene-liquid equilibria.

Negative polybaric fractional crystallization

For any given liquid composition a range of parental compositions can be generated by negative fractional crystallization. For EG4507, for example, parent liquids may lie some way back along the olivine-plagioclase cotectic and then even further back towards olivine alone. Alternatively at high pressure, parent liquids may lie along cotectics involving olivine+plagioclase+clinopyroxene, or olivine+clinopyroxene and ultimately olivine alone.

Two particular features emerge from these general considerations. (1) Provided that the same phases are involved at all stages, negative fractional crystallization of a range of compositions tends to produce convergence of their parental compositions. The small but significant differences in composition which would be amplified by subsequent positive fractional crystallization are more easily produced in converged parental compositions. For example, Ca/Al ratios are very sensitive to crystallization, or otherwise, of clinopyroxene. For this reason some of the range of Skaergaard chill compositions might be more easily related to each other by equilibria at very high pressure. (2) Polybaric fractional crystallization of a particular assemblage of phases from a particular liquid, accomplished by lowering pressure by a small selected amount after each increment of crystallization, may produce residual liquid compositions significantly different from the liquid compositions produced by isobaric fractional crystallization of the same phases. For example, positive isobaric fractional crystallization of olivine+plagioclase+clinopyroxene from EG4507 at 10 kbar produces residual liquids with Al_2O_3 contents rising to over 20% whereas polybaric fractional crystallization of these phases from 10 to 1 kbar produces residual liquids with Al_2O_3 contents falling to below 16%.

Conclusions

The integration of data from phase equilibrium experiments on a range of bulk rock compositions by means of a set of mineral-phase liquidus-temperature equations is sufficiently accurate to allow some quantitative modelling of crystal-liquid fractionation processes. The accuracy of this approach and the scope for extending its range of applicability to higher pressures and to equilibria involving additional phases such as garnet and amphibole are being examined. This approach should be able to

test for possible relationships between associated igneous rocks; to determine the conditions of formation of phenocryst assemblages; and to constrain cumulus sequence parental liquid compositions in cases for which no appropriate chilled liquid

sample is available. It is also easy to incorporate trace elements into the fractionation scheme.

I thank Drs K. G. Cox and G. M. Biggar for valuable comments and D. G. Russell for his unpublished glass analyses.

Received 31 July 1980; accepted 6 February 1981.

1. Wager, L. R. & Brown, G. M. *Layered Igneous Rocks* (Oliver & Boyd, Edinburgh, 1967).
2. McBirney, A. R. *Nature* **253**, 691–694 (1975).
3. Irvine, T. N. in *The Evolution of the Igneous Rocks, 50th Anniversary Perspectives* (ed. Yoder, H. S. Jr) 245–306 (Princeton University Press, 1979).
5. Blank, H. R. & Gettings, M. E. *Trans. Am. Geophys. Un.* **54**, 507 (1973).
6. Hoover, J. D. *Yb. Carnegie Instn Wash.* **77**, 732–743 (1978).
7. Maaløe, S. J. *Petrol.* **17**, 398–419 (1976).
8. Nwe, Y. Y. *Contr. Miner. Petrol.* **55**, 105–126 (1976).
9. Biggar, G. M. *Contr. Miner. Petrol.* **46**, 159–167 (1974).
10. Roedder, P. L. & Emslie, R. F. *Contr. Miner. Petrol.* **29**, 275–289.
11. Nathan, H. D. & Van Kirk, C. K. *J. Petrol.* **19**, 66–94 (1978).
12. McBirney, A. R. & Noyes, R. M. *J. Petrol.* **20**, 487–554 (1979).

13. Burnham, C. W. in *The Evolution of the Igneous Rocks, 50th Anniversary Perspectives* (ed. Yoder, H. S. Jr) 428–439 (Princeton University Press, 1979).
14. Pearce, T. M. *Nature* **276**, 771–774 (1978).
15. O'Hara, M. J. *Earth. Sci. Rev.* **4**, 69–133 (1968).
16. Wright, T. L. & Doherty, P. C. *Bull. geol. Soc. Am.* **81**, 1995–2008 (1970).
17. Hostetler, C. J. & Drake, M. J. *J. geophys. Res.* **85**, 3789–3796 (1980).
18. Russell, D. G. *Progr. exp. Petrol.* (in the press).
19. Ford, C. E. & Russell, D. G. (in preparation).
20. Ford, C. E. *Progr. exp. Petrol.* (in the press).
21. Bender, J. F. *et al. Earth planet. Sci. Lett.* **41**, 277–302 (1978).
22. Bickle, M. J. *et al. Earth planet. Sci. Lett.* **37**, 97–106 (1977).
23. Thompson, R. N. *Contr. Miner. Petrol.* **45**, 317–341, 1974.
24. Ford, C. E. *et al. Geochim. cosmochim. Acta Suppl.* **3**, 1, 207–229 (1973).
25. Fisk, M. R. *et al. Contr. Miner. Petrol.* **74**, 361–374 (1980).

Chloroplast protein phosphorylation couples plastoquinone redox state to distribution of excitation energy between photosystems

John F. Allen^{*†}, John Bennett[†], Katherine E. Steinback^{*‡} & Charles J. Arntzen^{*‡}

^{*} USDA-SEA-AR, Department of Botany, University of Illinois, Urbana, Illinois 61801, USA

[†] Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK

[‡] MSU/DOE Plant Research Laboratory, Michigan State University, East Lansing, Michigan 48824, USA

In photosynthetic membranes isolated from pea leaves, the redox state of the plastoquinone pool controls both the level of phosphorylation of the chloroplast light-harvesting pigment-protein complex (LHC) and distribution of absorbed excitation energy between the two photosystems. Phosphorylation of LHC polypeptides is proposed as the regulatory mechanism by which photosynthetic systems adapt to changing wavelengths of light.

In chloroplasts, non-cyclic photosynthetic electron transport generates ATP and NADPH by the operation of photosystem I and photosystem II acting in series. The energy that drives this process is derived partly from light absorbed by antenna chlorophylls associated closely with each photosynthetic reaction centre and partly from the more peripheral light-harvesting pigments¹ which are organized into specific chlorophyll-protein complexes^{2–4}. The most abundant of the chlorophyll-binding proteins in green plants is a 25,000 molecular weight (MW) polypeptide known as the light-harvesting chlorophyll *a/b* protein (LHCP).

In non-cyclic electron transport, the two photosystems must operate at the same rate. If absorbed radiant energy is distributed between the photosystems in such a way that one photosystem is excited more than the other, the excess excitation energy is dissipated, with a consequent loss of photosynthetic quantum efficiency. However, it is clear from studies with algal cells exposed to red and far-red light that chloroplasts can detect such an imbalance in the excitation of the photosystems and at least partially correct it by increasing the proportion of absorbed excitation energy that is distributed to the rate-limiting photosystem^{5–8}. These adaptive changes are referred to as state 1–state 2 transitions. State 1 is the adaptive state which results from over-excitation of photosystem I (as, for example, by far-red light) and is characterized by an increase in the proportion of total absorbed excitation energy that is transferred to photosystem II relative to photosystem I. State 2 results from over-excitation of photosystem II by red light and is characterized by increased energy transfer to photosystem I.

The biochemical events underlying state 1–state 2 transitions are poorly understood. However, a related phenomenon that has been extensively studied at the biochemical level is the ability of metal cations to affect the distribution of absorbed

excitation energy in isolated thylakoids^{9–11}. In the presence of salts of monovalent and divalent metal cations¹², the light-harvesting chlorophyll-protein complexes that contain LHCP are structurally and functionally associated with photosystem II, whereas in the absence of salts the association between the light-harvesting apparatus and the photosystems is altered and excitation energy transfer to photosystem I increases.

A major factor governing the response of thylakoids to cations *in vitro* is the LHC itself. Experiments using thylakoids have shown that trypsin can selectively remove a surface-exposed segment of the LHCP and that this removal is concomitant with the loss of the ability of cations to affect the distribution of excitation energy¹³. In addition, antibodies raised against isolated, purified LHC were found to react with surface-exposed segments of the membrane-bound complex and to prevent such cation effects¹⁴. These studies indicate that the regulatory mechanism controlling the distribution of absorbed excitation energy in chloroplasts involves a surface-exposed segment of the LHCP.

It has recently been established that the LHCP is reversibly phosphorylated on a threonine residue located within this surface-exposed segment^{15–18}. The LHCP belongs to a group of thylakoid polypeptides shown to be phosphorylated by a membrane-bound, light-activated protein kinase^{19,20} and dephosphorylated by a membrane-bound phosphatase¹⁷. Bennett *et al.*²¹ have demonstrated that the reversible phosphorylation of LHCP changes the light-harvesting properties of isolated thylakoids: phosphorylation increases (and dephosphorylation decreases) the proportion of absorbed excitation energy transferred to photosystem I.

This article shows that the activity of the protein kinase is regulated by the redox state of plastoquinone. We propose a molecular mechanism for state 1–state 2 transitions *in vivo* that

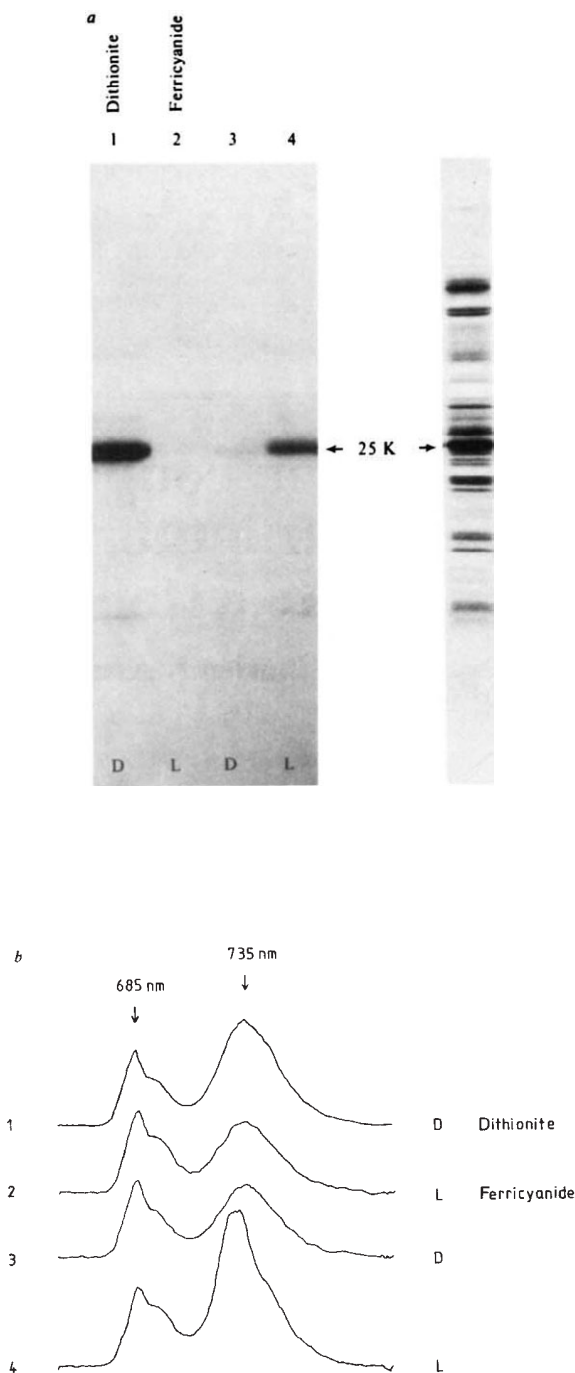


Fig. 1 Effects of light, dithionite and ferricyanide on phosphorylation of LHCP (a) and on 77 K fluorescence emission spectrum of thylakoids (b). a, Thylakoids that had been pre-equilibrated with or without dithionite (10 mM) in darkness and with or without ferricyanide (2 mM) in the light were incubated with [γ - 32 P]ATP as described in Table 1 legend. Total 32 P incorporation into acid-insoluble material was determined and then each sample was subjected to SDS-polyacrylamide gel electrophoresis, followed by staining (lane 5) and autoradiography (lanes 1–4). The identities of the lanes (together with total protein kinase activity of the thylakoids, expressed as a percentage of the light control) are: lane 1, dark + dithionite (124%); lane 2, light + ferricyanide (33%); lane 3, dark (16%); lane 4, light (100%). The position of the 25,000 MW LHCP is indicated (25 K). b, Thylakoids were pre-equilibrated and incubated as described above except that the dithionite concentration was 20 mM and the ATP was not radioactive. After incubation, thylakoids were quickly diluted 10-fold with a medium containing 50 mM Tricine-NaOH (pH 7.6), 30% (v/v) glycerol and 10 mM MgCl₂ and then frozen to 77 K. Fluorescence emission spectra were obtained as described previously²¹ using 440 $\pm$ 10 nm light for excitation. Spectra were normalized to give equal relative fluorescence at 685 nm using a Hewlett-Packard 9825 computer.

is consistent with these data and we argue that the presence of cations is merely a prerequisite for these transitions rather than a decisive regulatory factor in itself.

Activation of the thylakoid protein kinase by light and reducing agents

The protein kinase activity of isolated thylakoids was assayed by measuring the transfer of 32 P from [γ - 32 P]ATP to total membrane protein (Table 1). Light stimulated the kinase about fourfold compared with a dark-incubated control. The herbicide 3-(3',4'-dichlorophenyl)-1,1-dimethylurea (DCMU), which blocks electron transport between the primary acceptor (Q) of photosystem II and plastoquinone, completely abolished the stimulatory effect of light. The stimulation could not be restored by addition of 2,6-dichlorophenolindophenol (DCPIP) in the reduced form (DCPIPH₂), which donates electrons to the intersystem chain after plastoquinone. Thus, the kinase seemed to be stimulated by reduction of certain intersystem electron carriers. This idea was supported by the effects of oxidizing and reducing agents on kinase activity (Table 1). Ferricyanide and methyl viologen accept electrons from photosystem I and consequently tend to oxidize the intersystem carriers. These oxidants inhibited light-activation of the kinase. In contrast, the low-potential reductant dithionite²² activated the kinase in the dark, as did tetramethylhydroquinone (TMQH₂, a specific donor to plastoquinone in the intersystem electron transport chain^{23,39}). Although dithionite was a somewhat more effective activator than light, their effects were equivalent rather than additive or synergistic. It is likely, therefore, that light and dithionite activated the kinase by the same basic mechanism, that is, by reduction of the intersystem carriers.

When activated by light, the protein kinase of the thylakoid phosphorylates predominantly the 25,000 MW LHCP¹⁶. Figure 1a shows that the same protein was phosphorylated when the kinase was activated in the dark by dithionite. Dithionite also stimulated phosphorylation of the 8,000 MW polypeptide reported to be the proton channel of the thylakoid ATP

Table 1 Activation of thylakoid protein kinase by light and reducing agents

Incubation conditions	Kinase activity (%)
Light	100
Dark	28
Light + DCMU	21
Light + DCMU + DCPIPH ₂	18
Dark + dithionite	111
Light + dithionite	112
Light + DCMU + dithionite	121
Dark + TMQH ₂	101
Light + methyl viologen	27
Light + ferricyanide	17

Thylakoid membranes were prepared from 12–20-day-old peas (*Pisum sativum* L. var. Progress no. 9) grown as previously described³. Shoots (25 g) were homogenized for 5 s in a blender at low speed in 100 ml of 50 mM Tricine-NaOH (pH 7.8), 0.4 M sorbitol and 5 mM MgCl₂. The homogenate was filtered through cheesecloth, and centrifuged at 1,000g for 5 min. The pelleted chloroplasts were osmotically shocked in 50 ml of 10 mM Tricine-NaOH (pH 7.8) and 5 mM MgCl₂. The suspension was centrifuged at 2,800g for 5 min and the thylakoid pellet was resuspended to a final chlorophyll concentration of 1 mg ml⁻¹ in 50 mM Tricine-NaOH (pH 7.8), 100 mM sorbitol and 10 mM MgCl₂. Portions (100 μ l) of the thylakoid suspension were equilibrated at 20 °C in a final volume of 1 ml of the same buffer containing the indicated additions: 3-(3',4'-dichlorophenyl)-1,1-dimethylurea (DCMU) at 10 μ M; 2,6-dichlorophenolindophenol (DCPIP) at 50 μ M, maintained in the reduced form (DCPIPH₂) by sodium ascorbate (2 mM); sodium dithionite (20 mM); tetramethylhydroquinone (TMQH₂, 500 μ M) reduced with sodium borohydride as described elsewhere²³; methyl viologen (50 μ M); potassium ferricyanide (2 mM). Equilibration took place either in the light (15 μ E m⁻² s⁻¹) or darkness, and, where dithionite was included, reaction mixtures were made anaerobic by bubbling them with nitrogen gas before the addition of thylakoids. Phosphorylation reactions were started by addition of [γ - 32 P]ATP (10 Ci mol⁻¹) to a final concentration of 100 μ M. The reaction was stopped after 4 min by addition of 0.5 ml of cold trichloroacetic acid (30%, w/v). Acid-stable, precipitable radioactivity was counted by Cerenkov spectrometry. The extent of protein phosphorylation was calculated from the mean activity of duplicated samples. Results are expressed as a percentage of the rate obtained in the light in aerobic conditions (5 nmol phosphate incorporated per mg chlorophyll in 4 min).

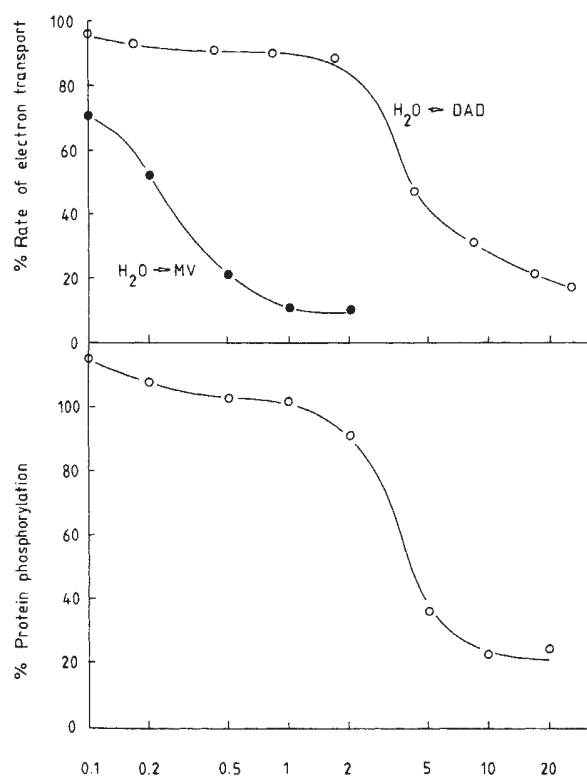


Fig. 2 Inhibition of electron transport and protein kinase activity by 2,5-dibromo-3-methyl-6-isopropyl-*p*-benzoquinone (DBMIB). Whole chain electron transport was measured as oxygen uptake in the presence of methyl viologen (MV, 100 μ M), sodium azide (5 mM) and ATP (100 μ M). The control rate in the absence of DBMIB was 23 μ mol (mg chlorophyll h)⁻¹. Photosystem II-mediated electron transport was measured as oxygen evolution in the presence of K₃Fe(CN)₆ (2.5 mM), diaminodurene (DAD, 1.2 mM) and ATP (100 μ M). The control rate was 97 μ mol (mg chlorophyll h)⁻¹. All electron transport measurements were carried out at saturating light intensity (1 mE m⁻²s⁻¹) in a reaction medium (2 ml) containing 50 mM Tricine (pH 7.8), 100 mM sorbitol, 10 mM MgCl₂ and chloroplasts equivalent to 200 μ g chlorophyll. Protein kinase activity was measured as described in Table 1 legend. DBMIB was present during the equilibration period before the addition of [γ -³²P]ATP. Values for phosphorylation were calculated by subtraction of the activity of a dark control and are expressed as a percentage of the activity in the light in the absence of DBMIB.

synthase²⁰. In contrast, ferricyanide completely inhibited phosphorylation of these thylakoid polypeptides.

Changes in the distribution of absorbed excitation energy

Bennett *et al.*²¹ reported that light-induced protein phosphorylation altered the 77 K chlorophyll fluorescence emission spectrum of isolated thylakoids. The finding that emission at 735 nm (F_{735} ; derived mainly from the chlorophylls of photosystem I) was increased relative to emission at 685 nm (F_{685} ; derived mainly from light-harvesting chlorophylls associated with photosystem II) was one of several lines of evidence²¹ indicating that protein phosphorylation changed the distribution of excitation energy within the pigment bed of the membrane in such a way as to favour distribution to photosystem I at the expense of photosystem II.

If the above conclusion is correct, changes in the distribution of absorbed excitation energy should occur whenever the level of protein phosphorylation is altered regardless of how the latter is brought about. Figure 1b shows that the 77 K chlorophyll fluorescence emission spectrum of isolated thylakoids was changed by ferricyanide and dithionite in accordance with the effects of these reagents on protein phosphorylation. Ferricyanide inhibited the light-dependent increase in F_{735} while

dithionite increased F_{735} in darkness, thus providing evidence for the control of excitation energy distribution by the redox state of the electron transport chain.

Although the dithionite-induced fluorescence change occurred in the presence of ATP, the data in Fig. 1 do not in themselves indicate that the change required ATP or involved protein phosphorylation. To explore this point in more detail, thylakoids were incubated in various conditions in the presence and absence of ATP and dithionite. Values for the ratio F_{735}/F_{685} were calculated from the resulting 77 K fluorescence emission spectra (Table 2). An ATP-dependent increase in the ratio was observed only in conditions which also permitted protein phosphorylation. Thus, there was no significant effect of ATP on fluorescence in the dark or in the light in the presence of DCMU, unless dithionite was also present. It is also clear from Table 2 that dithionite exerted two effects on the F_{735}/F_{685} ratio, one depending on the presence of ATP and the other not. The effect which required both ATP and dithionite represented the change in the distribution of excitation energy that could be attributed to dithionite-activated protein phosphorylation. The effect of dithionite in the absence of ATP may result from a direct alteration in fluorescence yield.

Activation of protein kinase by reduced plastoquinone

The above data raised the possibility that plastoquinone was the electron transport component whose reduction led to kinase activation. This possibility was tested by the use of 2,5-dibromo-3-methyl-6-isopropyl-*p*-benzoquinone (DBMIB), which inhibits electron transport at two sites—it inhibits oxidation of plastoquinone at a high-affinity site²⁴ and reduction of plastoquinone at a secondary, low-affinity site²⁵. Figure 2 illustrates both inhibitory activities. At a concentration of 1 μ M,

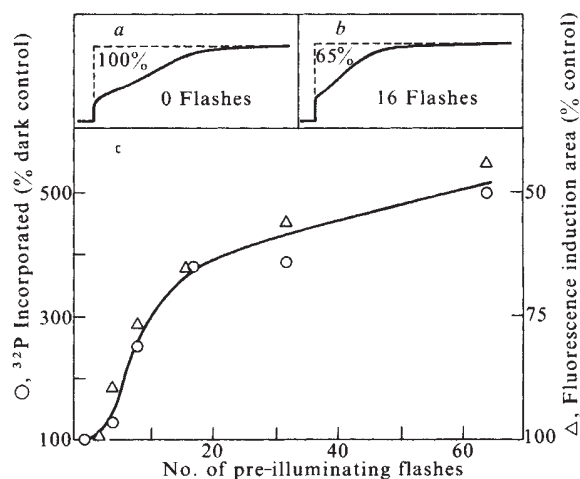


Fig. 3 Thylakoid protein phosphorylation and redox state of plastoquinone as functions of the number of pre-illuminating single-turnover flashes. Thylakoids were diluted with a medium containing 50 mM Tricine (pH 7.8), 100 mM sorbitol, 10 mM MgCl₂ and 10 mM NaF to a chlorophyll concentration of 50 μ g ml⁻¹ for protein kinase assays or 5 μ g ml⁻¹ for room temperature chlorophyll fluorescence induction transients. The dilution medium had been bubbled with nitrogen before the addition of thylakoids. The samples were pre-illuminated with a flashlamp (General Radio 'Stroboslav', type 1539A) delivering one intense, 8- μ s flash per s. A fibre-optics bundle carried light from the flash-lamp directly into either cuvettes (for the induction transients) or polyethylene microcentrifuge tubes (for kinase assays). After the indicated number of flashes, fluorescence induction was measured in the cuvettes as previously described³. *a, b*, Direct tracings of oscillographic recordings; broken lines delimit the area above the induction curve, a measure of the proportion of the plastoquinone pool that is in the oxidized state. *c*, [γ -³²P]ATP was added to the microcentrifuge tubes for a 2-min, dark, anaerobic incubation which was terminated by addition of trichloroacetic acid to a concentration of 10%. Protein phosphorylation was then measured as described in Table 1 legend. Fluorescence measurements and additions of [γ -³²P]ATP were made within 5 s of the final pre-illuminating flash.

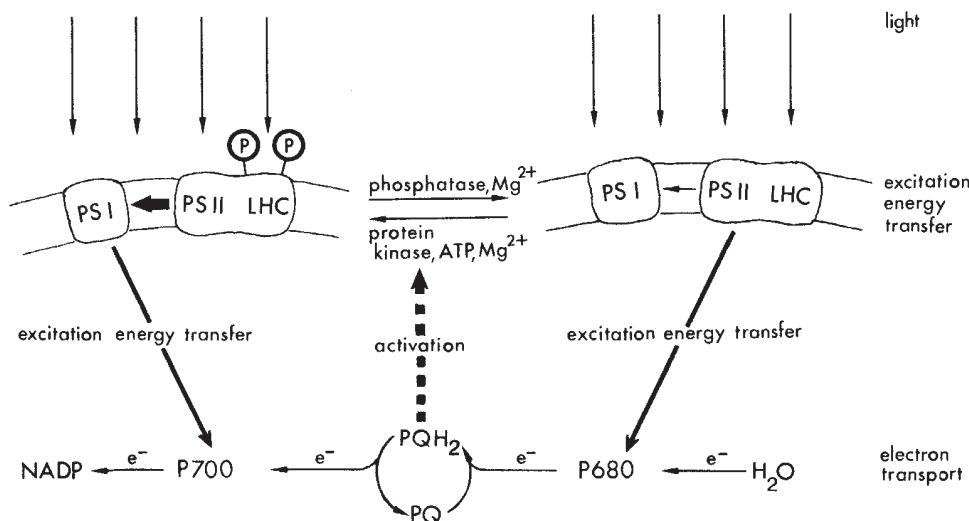


Fig. 4 A model for control of distribution of excitation energy in photosynthesis, in which reversible phosphorylation of LHC couples the redox state of plastoquinone (PQ) to the distribution of excitation energy between the photosystems. The mechanism we propose ensures maximal quantum efficiency of non-cyclic electron transport from water to NADP, mediated by P680 and P700, the reaction centres of photosystem II (PSII) and photosystem I (PSI), respectively. Reduction of plastoquinone by photosystem II leads to kinase activation and LHC phosphorylation, and hence to increased excitation of photosystem I. Conversely, oxidation of plastoquinone (PQH₂) by photosystem I inactivates the kinase; the phosphatase then dephosphorylates LHC and hence excitation of photosystem II is increased relative to that of photosystem I. Since the redox state of plastoquinone will be determined in part by the distribution of excitation energy between the two photosystems, reversible LHC phosphorylation completes a feedback loop by means of which any imbalance in this distribution will tend to be self-correcting.

DBMIB inhibited whole chain electron transport (from water to methyl viologen) by 90%, whereas photosystem II-mediated reduction of diaminodurene (DAD) was strongly inhibited only at concentrations greater than about 10 μ M (DAD accepts electrons at or before plastoquinone²⁶). When light-dependent kinase activity was measured as a function of DBMIB concentration (Fig. 2), strong inhibition was obtained only at concentrations equal to or greater than 10 μ M. Thus, the electron transport component that links the activity of the protein kinase to the redox state of the intersystem carriers is situated between the primary and secondary sites of action of DBMIB; the most obvious candidate is plastoquinone itself.

To characterize further the possible relationship between the redox state of plastoquinone and kinase activation, we subjected a series of thylakoid samples to short, intense flashes of light. The redox state of the plastoquinone pool was assessed by recording chlorophyll fluorescence induction transients (Fig. 3). The area above the curve of the transient is proportional to the fraction of the plastoquinone pool that is in the oxidized state²⁷. As shown in Fig. 3a and b, pre-illuminating flashes decreased the area above the curve of the transient, indicating that the plastoquinone pool had accumulated reducing equivalents. In parallel measurements, [γ -³²P]ATP was injected in the dark into preflashed samples which were then incubated for 4 min to permit kinase activity assay. As shown in Fig 3c, there was a striking similarity between the extent of kinase activation and that of reduction of the plastoquinone pool, as measured by the decrease in the area above the curve of the transient. Thus, the activity of the kinase is linked to the redox state of plastoquinone.

The previous observation¹⁹ that NADPH and ferredoxin could together activate the protein kinase in darkness may now

be explained as an indirect effect, resulting from the ability of reduced ferredoxin to reduce plastoquinone^{28,29}. We do not believe that activation of the thylakoid-bound protein kinase is mediated by either the thioredoxin system³⁰ or the 'LEM' system³¹, both of which have been implicated in the control of stromal enzyme activity and require electron transport via components on the reducing side of photosystem I.

Regulation of photosynthetic efficiency by reversible protein phosphorylation

The involvement of plastoquinone in the activation of the thylakoid protein kinase is probably of fundamental importance in the regulation of photosynthesis, for it provides a physiological mechanism for an adaptation which will correct any imbalance in the distribution of excitation energy between the two photosystems (Fig. 4).

The operation of this mechanism may be illustrated by reference to state 1-state 2 transitions. In green algae and higher plants, that is, in photosynthetic organisms containing LHCP, exposure to red light (of wavelength <680 nm) will excite photosystem II to a greater extent than photosystem I, and thereby lead to reduction of plastoquinone. The protein kinase will then be activated, LHCP will become phosphorylated and an increased proportion of absorbed excitation energy will be transferred from the light-harvesting complex to photosystem I at the expense of photosystem II. Thus, the biochemical event underlying the transition to state 2 is phosphorylation of the LHCP. Conversely, under far-red light (of wavelength >680 nm), photosystem I will be excited to a greater extent than photosystem II, plastoquinone will be oxidized and the kinase will be inactivated. Any phosphorylated LHCP will be dephosphorylated by the thylakoid phosphatase and excitation energy transfer to photosystem I will be decreased. Thus, the biochemical event underlying the transition to state 1 is dephosphorylation of the LHCP.

It has often been suggested that state 1-state 2 transitions reflect changes in the cation content of the stroma surrounding the thylakoids^{9-11,32}, a suggestion based on the effect of cations on the distribution of excitation energy in isolated thylakoids⁹⁻¹¹. One problem with this idea is that state 2 transitions would have to be correlated with a massive loss of cations from the stroma (to either the thylakoid or the cytoplasm), but there is no reason to believe that the cation concentration of the stroma changes greatly when higher plants or algae are transferred from far-red light to red light or vice versa. However, it should not be assumed that cations are not involved in the

Table 2 Effects of light, dithionite, DCMU and ATP on the ratio of 77 K fluorescence emissions at 735 nm and 685 nm

Incubation conditions*	F_{735}/F_{685}		
	+ATP	-ATP	
	(a)	(b)	(a)/(b)†
Dark	1.00	0.95	1.05
Light	1.90	1.20	1.58
Dark + dithionite	1.45	1.10	1.32
Light + dithionite	1.85	1.30	1.42
Light + DCMU	0.85	0.90	0.94
Light + DCMU + dithionite	1.90	1.30	1.46

* As described in Fig. 1b; DCMU concentration, 10 μ M.

† Values greater than 1 indicate an ATP-dependent change in excitation energy distribution in favour of photosystem I.

regulation of excitation energy distribution *in vivo*. Two roles need to be emphasized. First, cations such as Mg^{2+} ions are required to establish the pattern of excitation energy distribution seen in dephosphorylated membranes, including the close structural and functional association between the LHCP complex and photosystem II^{10,12,33}. The second role for Mg^{2+} ions derives from the fact that these cations are cofactors of both the kinase and the phosphatase (Fig. 4)^{17,19,20}. Thus, the presence of cations should be thought of as a prerequisite for state 1-state 2 transitions rather than as a regulatory factor *per se*.

The mechanism depicted in Fig. 4 provides an explanation for the original results of Bonaventura and Myers⁵ and of Murata⁶ and also for more recent contributions to the study of state 1-state 2 transitions. Several authors have drawn attention to the involvement of the high energy state of the membrane in these transitions^{9,34,35}. This would certainly be required for the generation of ATP consumed in protein phosphorylation (transition to state 2). Duysens³⁶ suggested that the redox state of a component of the intersystem electron transport chain controls these transitions. Plastoquinone would clearly fulfil this role. Horton and Black^{37,38} have recently established state 1-state 2 transitions *in vitro*. The transition to state 2 required ATP and was controlled by a membrane-bound factor with the redox

potential of plastoquinone. The idea that protein phosphorylation might provide the connection between plastoquinone and ATP is fully in agreement with our present results.

It is important to recognize that state 1-state 2 transitions are also observed in organisms that lack LHCP. For example, Murata⁶ and Ried and Reinhardt⁸ have shown that these transitions occur in red algae, which contain phycobiliproteins instead of LHCP complexes. Ried and Reinhardt⁸ conclude from their studies that the transitions are controlled by the redox state of a component of the intersystem electron transport chain. It will be interesting to discover whether state 1-state 2 transitions in the red algae are brought about by a mechanism analogous to that shown in Fig. 4; that is, a mechanism in which reversible phosphorylation of the phycobiliproteins is controlled by the redox state of plastoquinone.

This study was supported in part by NSF grant PCM 77-18953 to C.J.A. and by SRC grant GR/B11660 to J.B. DBMIB was synthesized and provided by Dr Don Ort. We thank Mrs J. Watson for technical assistance.

Note added in proof: Another piece of evidence for the mechanism in Fig. 4 has now been obtained. Potentiometric redox titrations of plastoquinone, of LHCP phosphorylation and of ATP-dependent chlorophyll fluorescence quenching give identical results⁴⁰.

Received 22 December 1980; accepted 18 February 1981.

- Butler, W. A. *Rev. Pl. Physiol.* **29**, 345-378 (1978).
- Thornber, J. P., Markwell, J. P. & Reinman, S. *Photochem. Photobiol.* **29**, 1205-1216 (1979).
- Burke, J. J., Ditto, C. L. & Arntzen, C. J. *Archs Biochem. Biophys.* **187**, 252-263 (1978).
- Mullet, J. & Arntzen, C. J. *Biochim. biophys. Acta* **589**, 100-117 (1980).
- Bonaventura, C. & Myers, J. *Biochim. biophys. Acta* **189**, 366-386 (1969).
- Murata, N. *Biochim. biophys. Acta* **172**, 242-251 (1969).
- Williams, W. P. in *Primary Processes in Photosynthesis* (ed. Barber, J.) 99-147 (Elsevier, Amsterdam, 1977).
- Ried, A. & Reinhardt, B. *Biochim. biophys. Acta* **592**, 76-86 (1980).
- Bennoun, P. & Jupin, H. *Proc. 3rd Int. Congr. Photosynthesis* **1**, 163-169 (1975).
- Barber, J. *FEBS Lett.* **118**, 1-10 (1980).
- Vernotte, C., Briantais, J. M., Armond, P. & Arntzen, C. J. *Pl. Sci. Lett.* **4**, 115-123 (1975).
- Arntzen, C. J. & Ditto, C. L. *Biochim. biophys. Acta* **449**, 259-274 (1976).
- Steinback, K. E., Burke, J. J. & Arntzen, C. J. *Archs Biochem. Biophys.* **195**, 546-557 (1979).
- Steinback, K. E., Burke, J. J., Mullet, J. E. & Arntzen, C. J. in *Chloroplast Development* (eds Akoyunoglou, G. & Argyroudi-Akoyunoglou, J. H.) 389-400 (Elsevier, Amsterdam, 1978).
- Bennett, J. *Nature* **269**, 344-346 (1977).
- Bennett, J. *Eur. J. Biochem.* **99**, 133-137 (1979).
- Bennett, J. *Eur. J. Biochem.* **104**, 85-89 (1980).
- Beliveau, R. & Bellemare, G. *Biochem. biophys. Res. Commun.* **88**, 797-803 (1979).
- Bennett, J. *FEBS Lett.* **103**, 342-344 (1979).
- Alfonzo, R., Nelson, N. & Racker, E. *Pl. Physiol.* **65**, 730-734 (1980).
- Bennett, J., Steinback, K. E. & Arntzen, C. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5253-5257 (1980).
- Ikegami, I. & Katoh, S. *Archs Biochem. Biophys.* **150**, 9-14 (1972).
- Izawa, S. & Pan, R. L. *Biochem. biophys. Res. Commun.* **83**, 1171-1174 (1978).
- Hauska, G. in *Encyclopedia of Plant Physiology* Vol. 5 (eds Trebst, A. & Avron, M.) 253-267 (Springer, Berlin, 1977).
- Gould, J. M. & Izawa, S. *Eur. J. Biochem.* **37**, 185-192 (1973).
- Izawa, S. in *Encyclopedia of Plant Physiology*, Vol. 5 (eds Trebst, A. & Avron, M.) 268-282 (Springer, Berlin, 1977).
- Murata, N., Nishimura, M. & Takamiya, A. *Biochim. biophys. Acta* **120**, 23-33 (1966).
- Arnon, D. I. & Chain, R. K. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4961-4965 (1975).
- Mills, J. D., Mitchell, P. & Barber, J. *Photobiophys. Photobiophys.* **1**, 3-9 (1979).
- Wolosiuk, R. A. & Buchanan, B. B. *Nature* **266**, 565-567 (1977).
- Anderson, L. E. & Avron, M. *Pl. Physiol.* **57**, 209-213 (1976).
- Murata, N. *Biochim. biophys. Acta* **189**, 171-181 (1969).
- Butler, W. L. *Brookhaven Symp. Biol.* **28**, 338-346 (1976).
- Murata, N. & Sugahara, K. *Biochim. biophys. Acta* **189**, 182-192 (1969).
- Krause, G. H. *Biochim. biophys. Acta* **292**, 715-728 (1973).
- Duysens, L. N. M. *Biophys. J.* **12**, 858-863 (1972).
- Horton, P. & Black, M. T. *FEBS Lett.* **119**, 141-144 (1980).
- Horton, P. & Black, M. T. *Biochim. biophys. Acta* **635**, 53-62 (1981).
- White, C. C., Chain, R. K. & Malkin, R. *Biochim. biophys. Acta* **502**, 127-137 (1978).
- Horton, P., Allen, J. F., Black, M. T. & Bennett, J. *FEBS Lett.* (in the press).

IgG antibodies to phosphorylcholine exhibit more diversity than their IgM counterparts

Patricia J. Gearhart

Department of Embryology, Carnegie Institution of Washington, 115 W. University Parkway, Baltimore, Maryland 21210, USA

Nelson D. Johnson, Richard Douglas & Leroy Hood

Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

An amino acid sequence analysis of the N-terminal immunoglobulin heavy and light chain variable regions (V_H and V_L) from 16 hybridoma proteins which bind phosphorylcholine as well as the complete sequence analysis of 9 of these V_H regions is presented. There seem to be more V_H regions participating in the phosphorylcholine response than can be encoded directly by germ-line V_H gene segments. Moreover, the V regions from IgG antibodies are considerably more variable than those from their IgM counterparts. These observations raise the possibility that a somatic mechanism for V region diversification produces greater diversity in IgG than in IgM antibodies.

ONE goal in molecular immunology is to understand the mechanisms responsible for antibody gene diversification. Insight into antibody diversity has been obtained by comparing sequence information from both immunoglobulin genes and proteins¹⁻⁸, and our present level of understanding may be summarized as follows. (1) Antibody molecules are composed of

heavy (H) and light (L) chains, each of which contain variable (V) and constant (C) regions. (2) Comparative analyses of many variable regions have demonstrated that amino acid sequence variability is unevenly distributed throughout the V region. The antigen-binding site is composed of extremely variable sections termed hypervariable regions (three for each chain), whereas

Fig. 1 V_L regions of anti-phosphorylcholine antibodies. The hybridomas were made by the following procedure. BALB/cJ mice (Jackson Laboratory) were immunized intraperitoneally with 0.1 mg of phosphorylcholine-haemocyanin⁵³ in complete Freund's adjuvant, and 1–2 months later (4 days before cell fusion), injected intraperitoneally with 0.1 mg of phosphorylcholine-haemocyanin in complete Freund's adjuvant. The cells from individual spleens were fused to non-secreting SP2/0-Ag14 cells⁵⁴ according to Kennett *et al.*⁵⁵. The heavy chain class of anti-phosphorylcholine antibodies was determined by radioimmunoassay⁵², and antibody with the T15 idiotype was detected by binding to A/J anti-T15 serum as described⁵³. The HPCM7(1F8), HPCG10(1B6), HPCG13(2F5), HPCG9(2B4), HPCG14(2C5), HPCG15(1B10), HPCG17(24B5) and HPCG12(1D8) hybridomas were obtained from J. Hurwitz⁵⁶. The HPCM3(63b4) and HPCM4(64a5) hybridomas were obtained from B. Clevering and J. Davie, and were generated from adult mice which were suppressed for expression of the T15

idiotype by treatment with anti-T15 serum at birth. The numbers in parentheses refer to the nomenclature of the donor laboratory. The antibodies were purified and chain separated as previously described^{57,58}. The amino acid sequence analysis was carried out on a modified Beckman 890B sequencer as previously described⁵⁹. Amino acids are noted by the one-letter code of Dayhoff⁶⁰. HV1 designates the first hypervariable region which ends at residue 40 (ref. 35). All * proteins come from one mouse, and the + proteins from a second mouse. Other proteins each come from separate individuals. The V_L sequences are categorized according to the T15, M603 and M167 groups with the most typical protein of each group used as the prototype. A solid line indicates identity with the prototype sequence. The sequence of HPC16 is written assuming Asx residues are identical to prototype asparagine or aspartic acid residues. The sequences of HPC52, HPC126, HPC19, HPC104 and HPC16 are from ref. 61; H8, S107, T15 and M603 from ref. 22; S63, W3207 and M511 from ref. 62; and M167 is from ref. 63. Y5236 sequence is from J. Hubert (personal communication).

	10	20	30	HV1	Class	T15 idiotype	Source
HPCM1	E V K L V E S G G G L V Q P G G S L R L S C A T S G F T F S			D F Y M E W	IgM	+	hybridoma
HPCM2					IgM	+	"
HPCM3					IgM	+	"
HPCM4					IgM	+	"
HPCM5*					IgM	+	"
HPCM6*					IgM	+	"
HPCM7*					IgM	+	"
HPC16					IgM	+	"
HPC35					IgM	+	"
HPC52					IgM	+	"
HPC126					IgM	+	"
HPCG8*					IgG3	+	"
HPCG9†					IgG3	+	"
HPCG10†					IgG3	+	"
HPCG11†					IgG3	+	"
HPCG12					IgG3	+	"
HPCG13†					IgG1	+	"
HPCG14†					IgG1	+	"
HPCG15					IgG1	+	"
HPCG17					IgG1	+	"
M603					IgA	+	myeloma
W3207					IgA	+	"
M167					IgA	+	"
M511					IgA	+	"
S63					IgA	+	"
Y5236					IgA	+	"
H8					IgA	+	"
S107					IgA	+	"
T15					IgA	+	"

non-binding portions, termed the framework regions, are more highly conserved^{9–12}. The hypervariable regions fold to constitute the walls of the antigen-binding site. (3) The V regions fall into sets of closely related sequences termed groups (formerly called subgroups)^{13–15}. The minor variation within the V regions of a particular group has afforded special insight into mechanisms of diversity^{2,3,5–8,16,17}. (4) The antibody molecules are encoded by three unlinked families of genes—two light chain families, lambda (λ) and kappa (κ), and a heavy chain family. (5) The V region of the light chain (V_L) is encoded by two distinct gene segments—V_L and J_L (joining)¹⁸, and the V region of the

heavy chain (V_H) is encoded by three distinct gene segments—V_H, D (diversity) and J_H (ref. 8). The J_L gene segments encode only a small portion of the third hypervariable region of the light chain, whereas the D and J_H gene segments encode most of the third hypervariable region of the heavy chain. (6) There are at least 100 germ-line V_L and V_H gene segments^{17,19}, indicating that germ-line diversity is extensive. (7) There are four functional J_κ and J_H gene segments^{6–8,20}, and it seems that any J_κ gene segment may be joined with any V_κ gene segment⁵. Likewise, the combinatorial joining of V_H, D and J_H gene segments is an important mechanism for amplifying diversity in the third hypervariable region of the heavy chain³. (8) The site-specific recombination mechanism which joins the V_L and J_L, or the V_H, D and J_H gene segments, may join particular gene segments at different points in their sequences, thus generating hybrid codons as well as codon insertions or deletions. Thus, alternative joining points generate somatic variation at these junctions during the act of gene segment joining^{4,6–8}. (9) The V_L regions seem to exhibit somatic variation in hypervariable and framework regions beyond that which can be explained by the germ-line V gene repertoire and the junctional diversity of V–J joining. In the V_λ region, this diversity appears exclusively in the hypervariable regions¹, but in V_κ regions, diversity is seen in framework regions as well⁵.

To assess the patterns of variable region diversity to a simple hapten, we have used the hybridoma technique of Köhler and Milstein²¹ to obtain homogeneous antibodies from BALB/c mice immunized against phosphorylcholine. Antibodies binding to phosphorylcholine were studied for the following reasons. First, the variable regions of a closely related set of myeloma proteins and serum antibodies that bind phosphorylcholine have been sequenced^{15,22–24}, and these data can be compared with the sequences obtained from hybridoma antibodies. Second, anti-idiotypic sera raised against several of the myeloma proteins have revealed that the majority of anti-phosphorylcholine antibodies from BALB/c mice have the variable region idiotype of the myeloma protein T15 (refs 25, 26); this is denoted the T15 idiotype. A variety of interesting analyses have subsequently been carried out on the developmental expression²⁷, genetic mapping^{28,29} and regulation^{30–32} of this predominant idiotype.

Table 1 Affinity of antibodies for phosphorylcholine

Antibodies	Class	T15 idiotype	K _a × 10 ⁵ M ⁻¹
T15-V _H , T15-V _L groups			
HPCM2	IgM	+	3.5
HPCM5	IgM	+	2.4
HPCM7	IgM	+	3.2
HPCG8	IgG3	+	1.3
HPCG14	IgG1	+	4.4
T15	IgA	+	4.3
T15-V _H , M603-V _L groups			
HPCM3	IgM	–	0.34
HPCG15	IgG1	–	0.13
W3207	IgA	–	2.5
T15-V _H , M167-V _L groups			
HPCG10	IgG3	–	0.25
HPCG13	IgG1	–	1.8
HPCG17	IgG1	–	1.0
M167	IgA	–	1.9

The affinity of antibodies for phosphoryl[Me-¹⁴C]choline (NEN) was measured by equilibrium dialysis in a buffer containing 0.17 M boric acid, 0.03 M sodium borate, 0.15 M sodium chloride, pH 8.0. Equilibrium dialysis was performed at 25°C in 0.1-ml compartments at an antibody concentration of 1.5–3.5 mg ml⁻¹. Protein molarity was determined by absorbance at 280 nm using $E_{1\text{cm}}^{1\%} = 1.37$, and a molecular weight of 150,000 for IgG and IgA, and 170,000 for IgM. The data were plotted by the method of Scatchard⁶⁷ to calculate number of sites and by the Sips distribution function⁶⁸ to calculate K_a. A least-squares fit was used to obtain the best line. The Scatchard plots were linear, and the number of binding sites ranged from 1.7 to 2.0 for monomeric forms. Values of α, the heterogeneity index, were calculated by the Sips distribution function and ranged from 0.9 to 1.06, indicating homogeneity of binding sites. Binding constants were accurate to ±10%.

We report here the N-terminal amino acid sequences for the V_H and V_L regions of 16 hybridoma antibodies which bind phosphorylcholine. In addition, we have determined the complete V_H region sequences for nine of these hybridoma antibodies. The most striking observation is that the IgG antibodies exhibit far more diversity than their IgM counterparts. The data also suggest that somatic variation may be involved in V region diversification.

Hybridoma and myeloma immunoglobulins belong to the same V_L and V_H groups

The N-terminal amino acid sequences of 30 V_L regions from hybridoma and myeloma proteins binding phosphorylcholine are given in Fig. 1. Immunoglobulin V_L regions which differ by two or less amino acids over the first 23 residues are classified as members of a single group¹⁵. The data indicate that the 30 V_L regions fall into three groups. These groups are the same as those initially delineated by analysis of light chains from myeloma proteins and are known as the T15, M603 and M167 groups³³. Light chains derived from antibodies to phosphorylcholine generated in mouse strains other than BALB/c also fall into these same three V_L groups³⁴. Each V_L group has a prototype sequence that is repeated identically many times (Fig. 1). Within a group, each V sequence that differs from the prototype will be denoted a nonprototype sequence.

Of 15 V_L regions in the T15 group, 14 are identical and the 15th differs by a single residue substitution which can be explained by a single nucleotide change. Three of six V_L regions in the M603 group are identical and the other three differ by two, two and five residues, respectively. All these substitutions can be accounted for by single nucleotide changes. Four of nine V_L regions in the M167 group are identical and the others differ by a single residue. Three of these substitutions require two nucleotide changes (HPCG10, HPCG9 and HPC16). Seven of the 15 substitutions noted in these groups fall outside the first hypervariable region as defined by Kabat *et al.*³⁵. There seems to be more diversity in the M603 and M167 V_L groups than in the T15 group.

The N-terminal amino acid sequences of the V_H regions from 11 IgM hybridomas, 9 IgG hybridomas and 9 IgA myeloma proteins that bind phosphorylcholine are shown in Fig. 2. Immunoglobulin V_H regions which differ by two or less amino acids over their N-terminal 27 residues belong to a single V group¹⁵. As seen in Fig. 2, all 29 V_H regions fall within a single V_H group. Twenty-two of the V_H sequences are identical, and six of the remaining seven V_H regions differ by a single residue from the prototype T15 sequence. The remaining V_H region, HPCG15, differs by six residues. Eleven of 12 substitutions can

be explained by single nucleotide substitutions, and 8 of 12 substitutions lie outside the first hypervariable region.

The complete V_H region sequences from 10 hybridoma antibodies and 9 myeloma proteins are compared in Fig. 3. The hybridoma and myeloma V_H regions exhibit similar diversity patterns by several criteria. (1) Three hybridoma and four myeloma V_H regions are identical to the T15 prototype V_H sequence. (2) The variant hybridoma V_H sequences differ by 1–6 substitutions and the variant myelomas differ by 1–13 substitutions. (3) In both the hybridoma and myeloma V_H regions, insertions and deletions occur only in the third hypervariable region. (4) In the hybridoma V_H regions, 20 of 25 (80%) variant residues are found in the three hypervariable regions, and similarly, in the myeloma V_H regions, 21 of 28 (75%) variant residues are located in the three hypervariable regions. Since the hybridoma and myeloma V_H regions exhibit the same basic patterns of variation, we will consider both sets of data together in analysing diversity in the V_H segments (residues 1–101), the D segments (residues 102–108) and the J_H segments (109–125), which are encoded respectively, by the V_H , D and J_H gene segments⁸.

The V_H segments demonstrate diversity in both framework and hypervariable regions

Ten V_H segments are identical and nine others differ by one to eight substitutions (Fig. 3). Generally, the variant V segments differ by two or more widely scattered substitutions. Note that each variant V segment is distinct from all the others. Twenty-four variant residues are present at 16 different positions. Four positions show variation in more than one V_H region: position 56 has four different amino acid substitutions, position 55 has two different substitutions, position 28 has two identical substitutions and position 40 has three substitutions, two of which are identical. Twelve of 24 substitutions (50%) fall outside the two hypervariable regions contained in the V segment. Single base changes can account for 22 of 24 amino acid substitutions, and two substitutions require two base changes. Thus any diversification mechanisms must account for diversity outside as well as within hypervariable regions²².

D segment diversity is extensive

Figure 3 demonstrates that the D segment diversity is extensive and falls into three categories: isolated single base changes (for example, position 106 in M167 and 107 in H8 and HPCG8), substitutions, insertions or deletions at either end of the D segment (as in HPCG13, HPCG14, M603, M511 and M167), and blocks of amino acid substitutions (for example, W3207, HPCM4 and HPCM6). The single base substitutions could arise

Fig. 2 The V_H regions of anti-phosphorylcholine antibodies. See Fig. 1 legend for an explanation of the symbols. The sequences of HPC16, HPC35, HPC52 and HPC126 are from ref. 61; M603 is from ref. 64; M167 from ref. 65; T15, W3207, M511, S63, H8 and S107 are from ref. 22; and Y5236 is from J. Hubert (personal communication).

	10	20	30	HV1	Class	T15	Idiotype	Source
HPCM1	D I V M T Q S P T F L A V T A S K K V T I S C		T A S E S L Y S S K H K V H Y		IgM	+		hybridoma
HPCM2					IgM	+		"
HPCM5*					IgM	+		"
HPCM6*					IgM	-		"
HPCM7†					IgM	-		"
HPC 52					IgM	+		"
HPCG8*		E			IgG3	+		"
HPCG11†					IgG3	+		"
HPCG12					IgG3	+		"
HPCG14†					IgG1	-		"
S63					IgA	+		myeloma
Y5236					IgA	+		"
H8					IgA	+		"
S107					IgA	+		"
T15					IgA	+		"
HPCM3	D I V M T Q S P S S L S V S A G E K V T M S C		K S S Q S L L N S G N Q K N Y		IgM	-		hybridoma
HPCM4					IgM	-		"
HPC126					IgM	-		"
HPCG15	S	A		R T R	IgG1	-		"
W3207				D G	IgA	-		myeloma
M603		R		F	IgA	-		"
HPC19	D I V I T Q D E L S N P V T S G E S V S I S C		R S S K S L L Y K D G K T Y L		IgM	-		hybridoma
HPC104					IgM	-		"
HPC16		V			IgM	-		"
HPCG9†				S	IgG3	-		"
HPCG10†				Q	IgG3	-		"
HPCG13†	N				IgG1	-		"
HPCG17					IgG1	-		"
M511		K			IgA	-		myeloma
M167					IgA	-		"

from the presence of multiple germ-line D gene segments or from somatic variation. The insertions or deletions of codons associated with either end of the D segment as well as the junctional codon substitutions probably arise from the site-specific recombinational mechanism that joins the V_H and D as well as the D and J_H gene segments (Fig. 4). This diversity arises because the actual DNA splice points may occur at different sites on each of these three gene segments, in each case generating insertions, deletions and/or hybrid codons at the junctional boundaries⁶⁻⁸. Figures 3 and 4 suggest that there are probably multiple germ-line D gene segments and that variable boundaries of recombination may occur for any of the V region gene segments— V_H , D or J_H . The blocks of substitutions may arise from multiple germ-line D segments, from the splicing together of two germ-line D gene segments or from a variety of other diversification mechanisms^{4,17,36,37}.

The number and diversity of germ-line D gene segments are not known, as the D gene segments have only recently been isolated from germ-line DNA (P. Early and L. Hood, personal communication). As a result of D segment variability, the third hypervariable region of the heavy chain is the most diverse portion of the antibody V regions³⁸. This can be clearly seen in the V_H regions of immunoglobulins binding phosphorylcholine as well as those binding dextran³.

The J_H segments are encoded by one germ-line J_H gene segment

The J_H segments of mouse heavy chains are encoded by four different J_H gene segments^{8,20}. In the 16 V_H regions from immunoglobulins binding phosphorylcholine that can be compared, all seem to have J_H segments coded by the J_{H1} gene segment (Figs 3, 4). Five J_H segments (HPCM6, HPCM4, HPCG14, M603 and W3207) have N-terminal substitutions or deletions at position 108. As discussed before, this diversity can be explained by variability in the DNA splice point for joining the D and J_H gene segments (Fig. 4). In addition, two J_H segments (W3207 and M167) have additional substitutions when compared with the T15 sequence. At first glance, these data suggest that a particular J_H gene segment may be selected by antibody specificity. However, this selection is not always the rule, for in the 19 V_H regions derived from hybridoma antibodies binding dextran, all four J_H gene segments are expressed³ (J. Scilling, personal communication).

Antibodies binding phosphorylcholine demonstrate combinatorial association between light and heavy chains

Among the IgM immunoglobulins binding phosphorylcholine, some use the same heavy chain in conjunction with two very different light chains (such as HPCM2 and HPCM3 in Figs 1 and 3). The T15- and M603-like V_L regions associated with these V_H regions differ by more than 50% of their amino acid sequence over their N-terminal 38 residues and belong to two different groups (Fig. 1). Moreover, this combinatorial association of two different light chains with the same heavy chain changes the nature of the antigen-binding site because the affinities of the HPCM2 and HPCM3 antibodies for phosphorylcholine differ by 10-fold (Table 1). It has long been assumed that one heavy chain may pair with many different light chains³⁹. Indeed, if 1,000 light and 1,000 heavy chains could be associated combinatorially, 10^6 different antibodies could be produced. Although it has not been formally demonstrated that every light chain can pair with every heavy chain, these data suggest that combinatorial association may occur in some antibodies to increase diversity.

The T15 idiotype is present on immunoglobulins with sequence heterogeneity in V_H and V_L regions

Antisera recognizing the T15 idiotype have been used to identify a restricted group of antibodies that bind to phosphorylcholine. The immunoglobulin class and idiotype classification of antibodies are given in Figs 1–3, and show that antibodies with the T15 idiotype are found among proteins of the IgM, IgG3 and IgA classes. These data also show that antibodies with the T15 idiotype can have different amino acid sequences in their V regions. Four hybridoma antibodies (HPCM1, HPCM2, HPCM5 and HPC52) and four myeloma proteins (S63, Y5236, S107 and T15) exhibit the T15 idiotype and have identical N-terminal V_H and V_L sequences. However, hybridomas HPCG8, HPCG11 and HPCG12 express the T15 idiotype but differ from the proteins described above by one or more residues either in the V_H (HPCG8, HPCG11 and HPCG12) or the V_L (HPCG8) region. These data clearly indicate that antibodies with the same idiotype comprise a

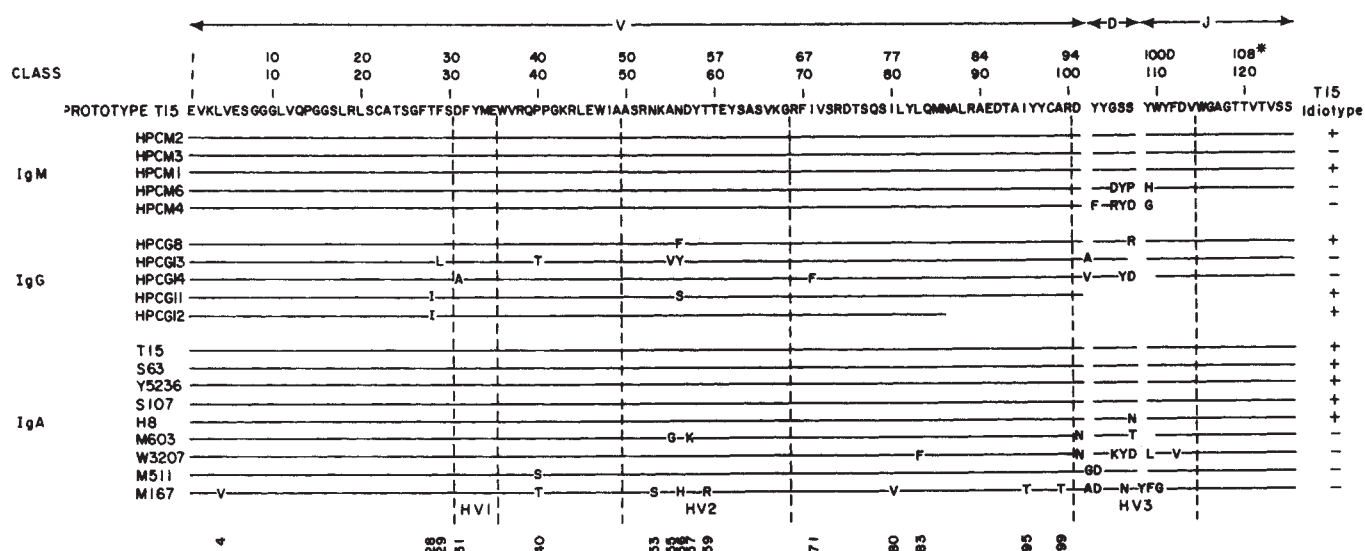


Fig. 3 Protein sequences of the V_H , D and J_H segments of anti-phosphorylcholine antibodies. The strategy for sequence analysis has been presented elsewhere⁶⁶. Hypervariable regions are indicated by vertical dotted lines (HV1, HV2 and HV3)³⁵ and the V, D and J segment boundaries are shown above the sequences. The numbering system of Kabat *et al.*³⁵ is shown on the first line (*) and is for M167. The consecutive numbering is shown on the second line and is for the M167 V_H region. All other V_H regions are shorter as indicated by the deletions at the V_H -D and D- J_H boundaries. Positions of amino acid substitutions are labelled at the bottom of the figure except in HV3. The references for data are given in Fig. 2 legend. The Asx residue reported at position 101 in M167 in ref. 35 has been changed to asparagine based on the DNA sequence⁸.

Fig. 4 Comparison of the DNA and protein sequences of three V_H regions at the V_H -D- J_H boundaries. The three sets of boxes indicate the distinct V_H , D and J_H gene segments. The DNA sequences are from ref. 8.

V_H										D					J_H				
DNA										TAC	TAC	GGT	AGT	A	GC	TAC	TGG	TAC	
M603										TAC	TAC	GGT	AGT	AC		C	TGG	TAC	
M167										G	GAC	TAC	GGT	AAT	AGC	TAC	TTT	G	
Protein										Tyr	Tyr	Gly	Ser	S		er	Tyr	Trp	Tyr
M603										Tyr	Tyr	Gly	Ser	Th			r	Trp	Tyr
M167										a	Asp	Tyr	Gly	Asn	Ser	Tyr	Phe	G	ly

family of closely related but distinct V regions, and, in this regard, are similar to immunoglobulins which bind nitrophenyl⁴⁰, dextran³ and arsonate^{41,42}.

Does somatic variation occur in the V_H and V_L segments derived from immunoglobulins binding phosphorylcholine?

The protein sequences presented here demonstrate that the V_H (amino acids 1–101) and V_L (amino acids 1–97) segments are very diverse. The question arises whether this V segment diversity can be generated entirely from multiple germ-line genes, or whether some somatic diversification must occur. We have approached this question by analysing the protein and nucleic acid data for the proteins and genes encoding the phosphorylcholine response.

To determine a minimal estimate of the number of V genes in the T15 V_H group, a cloned V_H probe was used to analyse undifferentiated (embryo) DNA by Southern blot analysis¹⁹. This V_H probe was derived from a full-length cDNA copy of the heavy chain mRNA from the S107 myeloma tumour⁴³. (The S107 and T15 V_H regions have identical protein sequences [Fig. 3].) The S107 V_H probe hybridizes strongly to four restriction fragments. Experiments have demonstrated that the S107 V_H probe cross-reacts strongly with mRNAs from M167 and M603⁴³. Because these are among the most different V_H regions from phosphorylcholine-binding immunoglobulins (Fig. 3), the S107 cDNA probe should detect most, if not all, of the V_H genes in the T15 group. If each of the DNA restriction fragments that hybridize to the S107 cDNA probe contains one T15-like V_H gene segment, there are approximately four V_H genes in the T15 V_H group. However, more than four different phosphorylcholine-specific myeloma and hybridoma proteins have already been sequenced. Among the nine myeloma V_H regions that are completely sequenced, four differ in their V segments from the T15 V_H region by one to eight residues (Fig. 3). Among the 20 hybridoma V_H regions that have been examined over their N-terminus, seven differ from the T15 V_H segment and from the myeloma V_H variants (Figs 2, 3). Thus, in the 29 V_H regions derived from immunoglobulins binding phosphorylcholine, 12 different V_H protein segments are present. The four germ-line DNA restriction fragments hybridizing to the S107 V_H probe may each contain more than one V gene segment. However, even if each band contained three V_H gene segments—the number necessary to explain the protein V_H segment data—one would expect, in 29 V_H segments analysed at the protein level, to see several examples of the same germ-line V_H gene being expressed two or three times. Instead, each nonprototype sequence is unique, and with two exceptions, every variant amino acid is different from other amino acids found at that position. Hence, we conclude that the pool of distinct V_H regions is very large and somatic variation must account for at least a portion of the nonprototype V_H sequences reported here. Interestingly, as judged by the data in Figs 2 and 3, variation occurs as frequently outside as within the hypervariable regions (16 of 31 V_H segment substitutions lie outside the hypervariable regions). Therefore, either the framework variants arise from distinct germ-line V_H gene segments or the presumptive mechanism for somatic variation is not confined to the hypervariable regions.

Somatic variation also may occur in anti-phosphorylcholine light chain genes. In the M167 V_L group, there is only one germ-line gene⁴⁴ (I. Weissman and U. Storb, personal communication). Six different protein sequences have already been identified (Fig. 1). Thus, somatic diversification also generates V_L variants.

An alternative explanation for this V region diversity is the possibility of genetic polymorphism in V gene segments among BALB/c mice. Although this possibility cannot be ruled out, it seems unlikely. BALB/c mice have been inbred for hundreds of generations, and mice obtained from a single source should be greater than 99% homozygous⁴⁵, although admittedly, some polymorphisms do exist in inbred strains⁴⁶. However, the extent of variation seen in the V_H and V_L regions from immunoglobulins binding phosphorylcholine far exceeds any known types of polymorphism including those noted in the globins and haptoglobins. For example, it is striking that four of the V_H and three of the V_L variants analysed in this study are derived from a single mouse (Figs 1, 2). Accordingly, individual mice are capable of generating extensive V segment diversity.

IgG antibodies binding phosphorylcholine exhibit more diversity than their IgM counterparts

The V_H segments from 11 IgM hybridomas do not show any diversity in their N-terminal 36 residues (Fig. 2). In contrast, the corresponding portions of the V_H regions from IgG hybridomas vary in six of the nine examples studied. The same pattern is seen in the corresponding V_L segments (Fig. 1). Of 12 light chains from IgM antibodies, 11 are identical in sequence to the prototype sequence of their respective groups while five of nine V_L regions from IgG antibodies vary. Thus, eight out of nine of the IgG antibodies vary from the V group prototype sequences in either their V_H or V_L regions, whereas only one of the 12 IgM antibodies varies.

One possible explanation for the observed differences in diversification between IgG and IgM proteins is that both do arise from the same population and the differences are the result of sampling errors. We used Fisher's exact method to demonstrate that it is very unlikely that these differences came from sampling error. The probability that the V_L segments of the IgM molecules and the V_L segments of the IgG molecules were selected from the same population is less than 0.05; likewise, the probability that the V_H segments from the IgM and IgG molecules were drawn from the same pool is 0.01; finally, the probability that both the V_L and the V_H segments were drawn from common V_L and V_H populations is less than 0.005. Thus, it is likely that the IgG polypeptides are significantly more diverse than their IgM counterparts. The complete V_H segment sequences strongly support this conclusion (Fig. 3) in that every V_H segment from the IgM group is identical to T15 and every V_H segment from the IgG shows one or more substitutions.

Cellular and recombinant DNA studies have suggested that B cells initially produce IgM and then switch to produce IgG or IgA after antigen stimulation^{20,47–49}. Thus, the greater diversity of the IgG hybridomas must arise either from infrequent non-prototype V_H gene segments in IgM-producing B cells, which are selectively expanded and expressed in IgG-producing cells, or from nonprototype V_H gene segments that are generated in

IgG-producing cells. Two important questions arise. First, how and when are the nonprototype sequences generated? Second, what selective pressures could give rise to the differences between IgM and IgG diversity?

The nonprototype sequences could be produced by one of two general mechanisms. (1) The nonprototype V regions may be encoded by germ-line V gene segments. We think this unlikely because there are apparently more protein sequences than germ-line genes in the T15 V_H group and M167 V_L group. (2) The nonprototype V regions may arise by somatic diversification. If so, B cells that express variant antibodies must be amplified by selection after the class switch from IgM to IgG production. Somatic diversification may occur by any one of three distinct mechanisms: ordinary somatic mutation³⁷, somatic recombination³⁹, or a special hypermutational mechanism^{36,50}. Two points are of special interest with regard to mechanisms of somatic mutation. First, activation of the somatic diversification mechanism may occur in conjunction with the switching of heavy chain classes. If so, the antibodies in secondary immune responses should be far more diverse than their primary immune response counterparts⁵¹. As the somatic diversification mechanism must generate variants in V_L as well as V_H gene segments, it must be activated in a *trans* fashion on different chromosomes. Second, somatic mutation may occur throughout the lifetime of individual B cells. Accordingly, the oldest B cells (those having undergone most clonal expansion) accumulate the most mutations. To explain our observations by this model, one would suggest that on the average B cells that had undergone a class switch were older than those which had not switched⁵².

Two types of selection may operate on infrequent germ-line or somatic variants which could result in the observed IgM-IgG differences. (1) Antigen-driven selection. B-cell clones expressing variant antibodies may be expanded because they have a higher affinity for antigen than B cells expressing prototype antibodies. As IgG-producing cell populations may have undergone more antigen-driven selection than IgM populations,

IgG antibodies should show more variability than IgM antibodies. At first sight, the data presented in Table 1 seem to argue against this possibility. The range of affinity constants seen in the IgM and IgG antibodies is not significantly different. However, these data do not exclude the possibility that the affinity constant of antibody for free hapten could be significantly different from the affinity constant of antibody for the hapten-carrier conjugate. Selection for variants in non-binding site regions may occur whenever framework residues modify the positions of binding-site residues. (2) Idiotype selection. Selection of infrequent variants may occur by idiotype-specific regulation. For example, if idiotype-specific suppression controls the level of predominant clones expressing the prototype sequences, B cells with altered sequences may be selectively expanded.

The IgA antibodies also are more diverse than their IgM counterparts ($P < 0.05$). However, the IgA V_H regions from immunoglobulins binding phosphorylcholine fall into two categories—those which have diversified (5/9) and those which are identical to the T15 prototype sequence (4/9) (Figs 1, 3). We propose that the V_H and V_L segments of IgG3, IgG1, IgA and probably IgG2a and IgG2b antibodies generally will be significantly more diverse than the V_H and V_L segments of IgM antibodies. Thus, increased diversity seems to be associated with the class switch from IgM to other isotypes. The increased variability in V segments is in part due to the somatic diversification of germ-line V genes. It will be interesting to compare the amino acid sequences of the protein variants with the nucleic acid sequences of the germ-line genes to determine which amino acid residues are encoded in the germ-line DNA and which result from somatic variation.

This work was supported by NIH grants CA 25507 and AI 10781. We thank Nadine Nivera for technical assistance, Fred Karush for help in performing equilibrium dialysis experiments, Julia Hurwitz and Joseph Davie for hybridomas, Doug Fambrough, Henry Huang, Steve Crews and Mitch Kronenberg for helpful discussions, Tim Hunkapiller for computer graphics, and J. Myron Hood for help with the statistical analyses.

Received 18 December 1980; accepted 27 February 1981.

- Weigert, M. & Riblet, R. *Cold Spring Harb. Symp. quant. Biol.* **41**, 837-846 (1976).
- Tonegawa, S., Hozumi, N., Mathysens, G. & Schuller, R. *Cold Spring Harb. Symp. quant. Biol.* **41**, 877-889 (1976).
- Schilling, J., Clevering, B., Davie, J. M. & Hood, L. *Nature* **283**, 35-40 (1980).
- Hood, L. *et al. Cold Spring Harb. Symp. quant. Biol.* (in the press).
- Weigert, M., Gattmaitan, L., Loh, E., Schilling, J. & Hood, L. *Nature* **276**, 785-790 (1978).
- Max, E., Seidman, J. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3450-3454 (1979).
- Sakano, H., Hüppi, K., Heinrich, G. & Tonegawa, S. *Nature* **280**, 288-294 (1979).
- Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. *Cell* **19**, 981-992 (1980).
- Wu, T. & Kabat, E. *J. exp. Med.* **132**, 211-250 (1970).
- Capra, J. D. & Kehoe, J. *Proc. natn. Acad. Sci. U.S.A.* **71**, 845-848 (1974).
- Amzel, L., Poljak, R., Saul, F., Varga, J. & Richard, F. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1427-1430 (1974).
- Padlan, E., Segal, D., Spande, T., Davies, D., Rudikoff, S. & Potter, M. *Nature new Biol.* **245**, 165-167 (1973).
- Hood, L., Gray, W., Sanders, B. & Dreyer, W. *Cold Spring Harb. Symp. quant. Biol.* **32**, 133-146 (1967).
- Milstein, C. *Nature* **216**, 330-332 (1967).
- Potter, M. *Adv. Immun.* **25**, 141-211 (1977).
- Bernard, O., Hozumi, N. & Tonegawa, S. *Cell* **15**, 1133-1144 (1978).
- Seidman, J., Leder, A., Nau, M., Norman, B. & Leder, P. *Science* **202**, 11-17 (1978).
- Brack, C., Hiram, M., Lenhard-Schuller, R. & Tonegawa, S. *Cell* **15**, 1-14 (1978).
- Davis, M., Early, P., Calame, K., Livant, D. & Hood, L. *ICN-UCLA Symposium on Eucaryotic Gene Regulation* (eds Axel, R., Maniatis, T. & Fox, C. F.) 393-406 (Academic, New York, 1979).
- Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. *Nature* **286**, 676-683 (1980).
- Köhler, G. & Milstein, C. *Nature* **256**, 495-497 (1975).
- Hood, L. *et al. Cold Spring Harb. Symp. quant. Biol.* **41**, 817-836 (1976).
- Klusens, L., Lee, W. & Köhler, H. *Eur. J. exp. Med.* **5**, 489-496 (1975).
- Rudikoff, S. & Claffin, J. L. *J. exp. Med.* **144**, 1294-1304 (1976).
- Sher, A. & Cohn, M. *Eur. J. Immun.* **2**, 319-326 (1972).
- Cosenza, H. & Köhler, H. *Science* **176**, 1027-1029 (1972).
- Sigal, N. H., Gearhart, P. J., Press, J. L. & Klinman, N. R. *Nature* **259**, 51-52 (1976).
- Lieberman, R., Potter, M., Mushinski, E. B., Humphrey, W. & Rudikoff, S. *J. exp. Med.* **139**, 983-1001 (1974).
- Gearhart, P. J. & Cebra, J. J. *Nature* **272**, 264-265 (1978).
- Strayer, D. S., Cosenza, H., Lee, W. M. F., Rowley, D. A. & Köhler, H. *Science* **186**, 640-643 (1974).
- Bottomly, K. & Mosier, D. E. *J. exp. Med.* **150**, 1399-1409 (1979).
- Gearhart, P. J. & Cebra, J. J. *J. exp. Med.* **149**, 216-227 (1979).
- Barstad, P. *et al. Science*, **183**, 962-964 (1974).
- Claffin, J. L. *Eur. J. Immun.* **6**, 669-674 (1976).
- Kabat, E. A., Wu, T. T. & Bilofsky, H. *Sequences of Immunoglobulin Chains* (US Department of Health, Education and Welfare, 1979).
- Baltimore, D. *Nature* **248**, 409-411 (1974).
- Cohn, M. *et al. ICN-UCLA Symposium on the Immune System: Genes, Receptors, Signals* (eds Sercarz, E., Williamson, A. R. & Fox, C. F.) 89-117 (Academic, New York, 1974).
- Cebra, J., Koo, P. & Ray, A. *Science* **186**, 263-265 (1974).
- Gally, J. A. & Edelman, G. A. *Rev. Genet.* **6**, 1-46 (1972).
- Reth, M., Hämmerling, G. J. & Rajewsky, K. *Eur. J. Immun.* **8**, 393-400 (1978).
- Marshak-Rothstein, A., Siekevitz, M., Margolies, M. N., Mudgett-Hunter, M. & Gelfer, M. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1120-1124 (1980).
- Estess, P., Lamoyi, E., Nisonoff, A. & Capra, J. D. *J. exp. Med.* **151**, 863-875 (1980).
- Early, P. W., Davis, M. M., Kaback, D. B., Davidson, N. & Hood, L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 857-861 (1979).
- Joho, R., Weissman, I. L., Early, P., Cole, J. & Hood, L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1106-1110 (1980).
- Klein, J. *Biology of the Mouse Histocompatibility-2 Complex* (Springer, New York, 1975).
- Potter, M. *Cell* **21**, 589-590 (1980).
- Raff, M. C., Megson, M., Owen, J. J. T. & Cooper, M. D. *Nature* **259**, 224-226 (1976).
- Davis, M. *et al. Nature* **283**, 733-739 (1980).
- Kataoka, T., Kawakami, T., Takahashi, N. & Honjo, T. *Proc. natn. Acad. Sci. U.S.A.* **77**, 919-923 (1980).
- Brenner, S. & Milstein, C. *Nature* **211**, 242-244 (1966).
- Rodwell, J. D. & Karush, F. *Molec. Immun.* **17**, 1553-1561 (1980).
- Karush, F. *Comprehensive Immunology*, Vol. 5 (eds Litman, G. & Good, R.) 85-116 (Plenum, New York, 1978).
- Gearhart, P. J., Sigal, N. H. & Klinman, N. R. *J. exp. Med.* **141**, 56-71 (1975).
- Shulman, M., Wilde, C. D. & Köhler, G. *Nature* **276**, 269-270 (1978).
- Kennett, R. H., Denis, K. A., Tung, A. S. & Klinman, N. R. *Curr. Topics Microbiol. Immun.* **81**, 77-91 (1978).
- Hurwitz, J., Coleclough, C. & Cebra, J. *Cell* **22**, 349-359 (1980).
- Cesebro, B. & Metzger, H. *Biochemistry* **11**, 766-771 (1972).
- Johnson, N. D., Hunkapiller, M. W. & Hood, L. E. *Analyt. Biochem.* **100**, 335-338 (1979).
- Hunkapiller, M. W. & Hood, L. E. *Biochemistry* **17**, 2124-2133 (1978).
- Dayhoff, M. O. *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Silver Spring, Maryland, 1976).
- Kocher, H. P., Berek, C., Schreier, M. H., Cosenza, H. & Jaton, J.-C. *Eur. J. Immun.* **10**, 264-267 (1980).
- Barstad, P. *et al. Eur. J. Immun.* **8**, 497-503 (1978).
- Rudikoff, S. & Potter, M. *Biochemistry* **17**, 2703-2707 (1978).
- Rudikoff, S. & Potter, M. *Biochemistry* **13**, 4033-4038 (1974).
- Rudikoff, S. & Potter, M. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2109-2112 (1976).
- Johnson, N. thesis, California Inst. Technol. (1980).
- Scatchard, G. *Ann. N.Y. Acad. Sci.* **51**, 660-672 (1949).
- Karush, F. *Adv. Immun.* **2**, 1-40 (1962).

Primary structure of a murine transplantation antigen

J. E. Coligan*, T. J. Kindt*, H. Uehara†, J. Martinko† & S. G. Nathenson†

*Laboratory of Immunogenetics, Building 8, Room 100, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20205, USA

†Departments of Microbiology and Immunology and Cell Biology, Albert Einstein College of Medicine, Bronx, New York 10461, USA

A radiochemical technique has been used to determine the primary structure of a murine H-2K^b molecule, a 346-residue glycoprotein subunit attached to a 99-residue β_2 -microglobulin subunit. Comparison of this structure with those available for other murine H-2 antigens and human HLA molecules suggests an evolutionary origin of MHC products.

THE antigens encoded within the tightly linked major histocompatibility complex (MHC)¹⁻⁴ are a diverse group of membrane-bound molecules intimately involved in various aspects of immune function. This genetic region, which has been characterized in several species, encodes a series of integral cell membrane glycoproteins with a high degree of polymorphism. Although the details of their function remain unknown, it is clear that MHC-encoded molecules are involved in the self-self recognition required for cellular cooperation^{1,2} and in the

recognition of altered cells by cytolytic T lymphocytes^{5,6}. An extensive list of traits, including susceptibility to specific diseases in humans, has also been linked to this complex locus^{7,8}. Although the MHC has been studied extensively by immunogenetic methods^{3,4}, only recently have details of primary structure been elucidated for human and murine MHC gene products^{9,10}.

Biochemical studies have shown that three major classes of molecules are encoded by the MHC, each having a different role

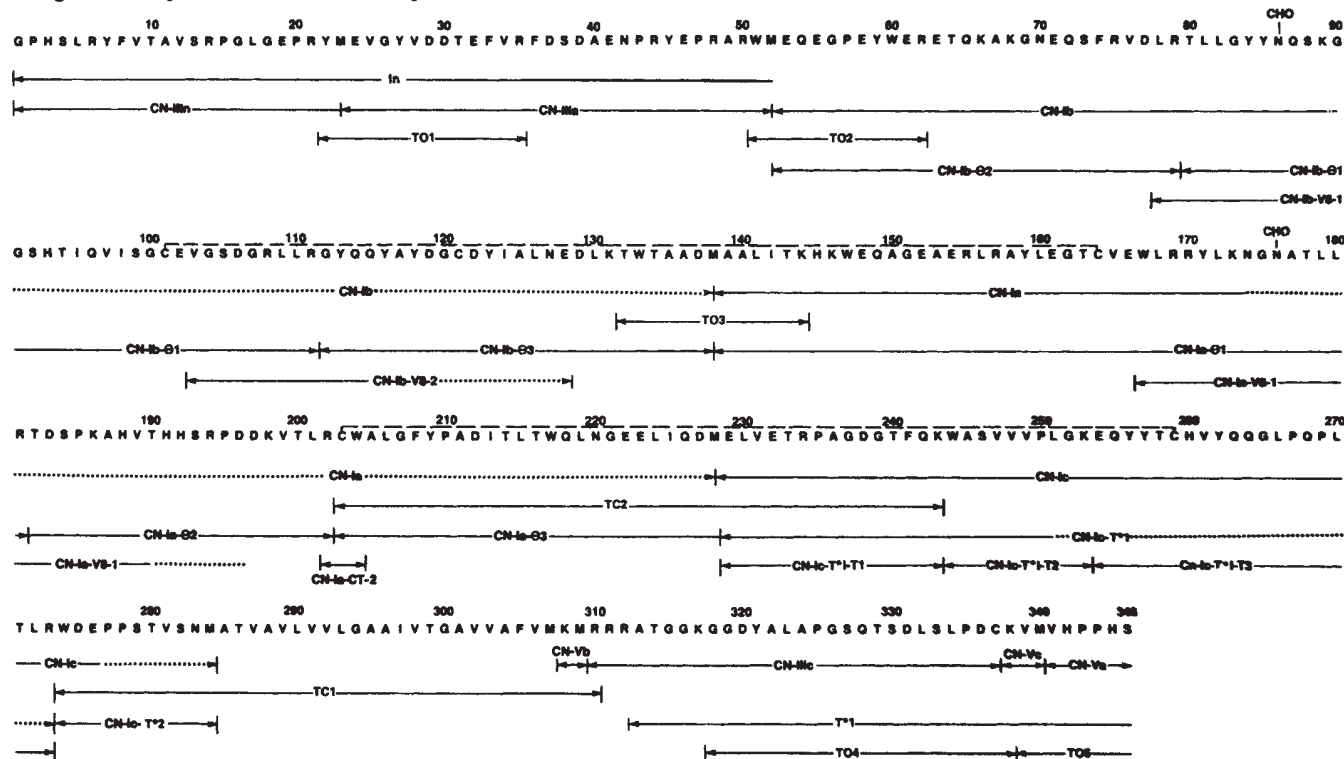


Fig. 1 Summary of information used to deduce the amino acid sequence of the H-2K^b glycoprotein^{12,13,20,21}. The sequence is given in single letter code where A is Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr. CHO indicates a carbohydrate moiety. Peptides are designated by letters indicating method of chain cleavage (In, intact molecule; CN, cyanogen bromide; TO, tryptic peptide obtained from the intact molecule used to overlap CNBr peptides; T*, trypsin after blockage of Lys with citraconyl groups; Θ, thrombin; V8, staphylococcal V8 protease; CT, chymotrypsin). TC refers to tryptic core, that is, tryptic peptides that were precipitated from the digestion mixture by 5% (v/v) acetic acid, collected by centrifugation, dissolved in 65% acetic acid and isolated by subsequent chromatographic procedures¹². When multiple designations are used, the order indicates the sequence in which the digestions were performed. For example, TO2 was obtained by tryptic cleavage of the intact chain whereas CN-Ib-V8-2 was produced by V8 treatment of the isolated cyanogen bromide fragment Ib. Fragment CN-Vc results from a reproducible but inappropriate CNBr cleavage. Dotted lines indicate peptide regions for which sequence data were incomplete. The strategy used for determining the complete sequence of the H-2K^b molecule relied on initial isolation of fragments obtained by CNBr cleavage (CN) at Met residues. Subsequent amino acid sequence analyses and use of tryptic overlap (TO) peptides permitted the alignment of these fragments. The only portion of the molecule not recovered as a CNBr fragment was the membrane-associated region (residues 284-307), which was obtained as a tryptic fragment from the intact molecule⁵⁶. The amino acid sequences of the small CNBr fragments IIIa, IIIb, Va, Vb and Vc were determined on the intact fragments. The largest CNBr fragments, Ia and Ib, were reduced to smaller fragments by thrombin cleavage in order to effect their sequence determination. The thrombin fragments were overlapped by isolating fragments obtained by staphylococcal V8 protease digestion and chymotryptic digestion of the appropriate CNBr fragments. Small peptides from CNBr fragment Ic were obtained by trypsinization after blocking of the Lys residues by citraconylation; after removal of the citraconyl groups, the large peptide was redigested with trypsin. Similarly, trypsin digestion after citraconylation was used to obtain fragments for overlapping the COOH-terminal CNBr fragments (IIIc, Vc, Va).

in the immune response^{7,11}. The class I molecules, which include the classical transplantation antigens, are composed of two noncovalently associated polypeptide chains, a 45,000-molecular weight polymorphic chain encoded by the MHC and a non-MHC-encoded 11,500-molecular weight polypeptide, β_2 -microglobulin (β_2 M)^{9,11}. Studies in our laboratories have concentrated on determining the primary structure of murine class I molecules¹²⁻¹⁸, and we report the complete primary

structure of one of these, H-2K^b, a molecule which has recently been intensively studied^{12,13,18-22}. The membrane-bound heavy chain of this glycoprotein is 346 amino acids long, and has two intra-chain disulphide linkages and two carbohydrate prosthetic groups linked to asparagine residues, whereas the associated β_2 M is 99 residues long, contains one intra-chain disulphide bond and is nonglycosylated. All our structural analyses of H-2K^b and several other murine class I antigens have been accomplished by the use of radiochemical methodology^{9,23}.

Isolation of radiolabelled H-2 alloantigens

There were two major obstacles to the study of the primary structure of H-2 alloantigens. First, because H-2 alloantigens are integral membrane proteins, they possess a hydrophobic region which associates with the membrane lipid bilayer, rendering them insoluble in common aqueous buffers, and thereby preventing the use of conventional biochemical purification techniques. This difficulty has been circumvented by using buffers containing a nonionic detergent (Nonidet P40) which solubilize H-2 antigens in high yield²⁴. Second, because only small amounts of H-2 alloantigens can be isolated from tissues²⁵ or from cultured cells^{26,27}, our primary structure determinations used H-2 alloantigens biosynthetically labelled with radioactive amino acids^{18,28,29}.

The procedure for obtaining radiolabelled H-2 antigens has been described previously^{14,30}. Briefly, appropriate cells are cultured in the presence of radioactive amino acids and the plasma membranes are then solubilized with Nonidet P40. The detergent-soluble extract is passed over a lentil lectin affinity column to obtain a glycoprotein fraction enriched in H-2 alloantigens. The H-2 molecules are isolated from this fraction by immune precipitation with monospecific alloantisera¹⁷ or monoclonal antibodies³¹. The H-2 heavy chain is dissociated and separated from the light chain (β_2 M) by gel filtration in 1 M formic acid.

For radiosequence studies, tumour cell lines are preferred as a source of material because they are metabolically more active than normal cells and thus incorporate greater amounts of the radiolabelled amino acids into cell protein; also tumour cell lines expressing increased amounts of the H-2 molecules under study can usually be obtained¹⁴.

In addition to the increased sensitivity gained in the detection of peptides and phenylthiohydantoin (PTH) amino acids, a major advantage of using radiolabelled material is that the H-2 molecules can then be purified rapidly by immune precipitation^{18,28-30} as there is no risk of the radiolabelled H-2 molecules being contaminated with unlabelled proteins (in fact, it is helpful to have additional protein to prevent losses). Also, because the original radiolabelled glycoprotein fraction remains essentially intact, several different molecules can be isolated from the same cell extract by sequential immunoprecipitation steps^{31,32}.

Sequence determination of radiolabelled proteins

The sequence of radioactive proteins and peptides is determined by stepwise degradation using the automated Beckman sequencer followed by analysis of PTH amino acid derivatives by HPLC, in much the same manner as used for sequencing unlabelled peptides. An obvious difference, however, is that instead of detecting the PTH amino acid derivatives directly on the HPLC, the radioactive PTH derivatives are analysed by co-chromatography with a mixture of unlabelled PTH amino acid derivatives. Peaks are collected and the radioactive amino acid residue is identified by liquid scintillation counting³³.

An important consideration in radiochemical sequence determination concerns the number of amino acids used in a given radiolabelled H-2 preparation. Although this can vary from 1 to 20, difficulties in detecting low specific activity amino acids in the presence of high specific activity amino acids make it practical to use groups of two to six amino acids^{14,21}. The

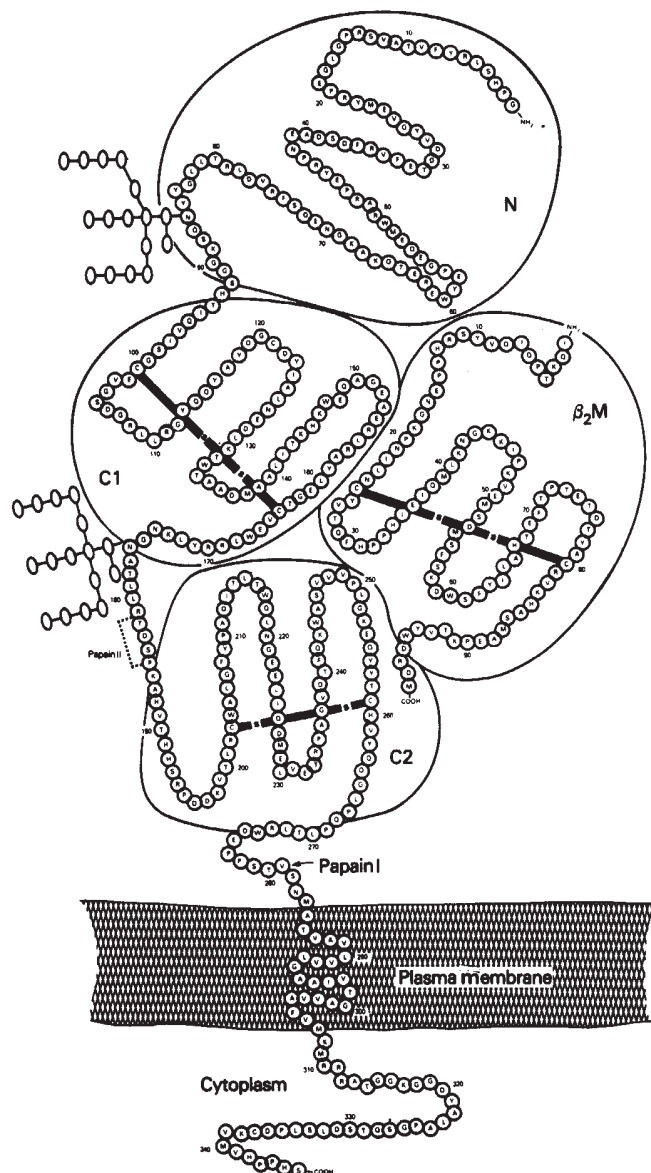


Fig. 2 A schematic representation of the complete amino acid sequence of the murine major transplantation antigen H-2K^b including the associated β_2 M and showing putative features of the molecule deduced from primary structure and other properties. The amino acids are identified within the circles in the single letter code detailed in Fig. 1 legend. The Cys-Cys bonds are depicted as black bars. The sequence shown for β_2 M is that for the variant found in C57BL/6 mice which has Ala at position 85 (ref. 22). The molecule may be considered to consist of three functional segments: the extracellular segment (~1-281), the membrane-bound segment (~282-308) and the cytoplasmic segment (~309-346). Cleavage by papain at the major site (papain I) yields a soluble antigenically active molecule lacking the hydrophobic and intracellular portions. Division of the extracellular portion of the molecule into three approximately equal subregions is based on homology of the CHO units 90 residues apart (N and C1) and the presence of two intra-chain disulphide linkages (C1 and C2). The globular nature of the N, C1 and C2 subregions is suggested by limited susceptibility of the molecule to papain digestion; the only fragments yielded are the intact extracellular segments and, in lower yield, a fragment designated papain II that includes the N and C1 subregions^{15,57,58}. The dotted box at positions 182-185 shows the approximate location of the papain cleavage yielding papain II. The β_2 M is not altered by papain treatment.

radioactive amino acids which become incorporated into cell protein at similar specific activities are grouped together. In determining such groups of amino acids and the amount of each to be incorporated, metabolic interconversions of certain amino acids must be considered^{14,33}. For example, radiolabelling cells with glutamine results in radioactive glutamine, glutamic acid and proline in the cell protein in an approximate ratio of 7:2:1 (ref. 20). This and other interconversions of amino acids have been consistently observed in two murine tumour cell lines^{14,33} and a rabbit cell line³⁴. Once known, the interconversions present no special difficulty. All amino acids except aspartic acid have been incorporated at sufficient levels for radiosequence analysis. Aspartic acid residues are assigned indirectly after concluding that a given position does not contain one of the other 19 possible amino acids^{12,13,21}. (Recent improvements in methods of radiolabelling cells will make possible the direct assignment of Asp residues³⁵.)

Structural features of H-2K^b

The primary structure of the H-2K^b glycoprotein was determined as outlined in Fig. 1. The primary structure of the β_2 M was investigated using radiolabelled protein separated from the H-2K^b, K^d and D^d heavy chains by gel chromatography in 1 M formic acid. Figure 2 depicts schematic representation for the structure of the H-2K^b transplantation antigen suggested from the amino acid sequence and other ancillary data. The 99-residue subunit (β_2 M) contains no carbohydrate unit, exhibits limited strain-related polymorphism^{22,36-38}, is noncovalently associated with the H-2-encoded chain, and shows no indication that it is bound to the cell membrane. The larger of the two chains contains two carbohydrate units and two intra-chain disulphide linkages and exhibits the high degree of polymorphism that characterizes H-2 alloantigens.

A major feature of the molecule is the ease with which the intact extracellular region can be cleaved from the cell surface with the proteolytic enzyme papain^{39,40}. Comparison of the structure of the papain-derived molecule (papain I) with that of the intact form has defined the COOH-terminus of the papain-derived molecule as residue 281 (ref. 13). This site coincides with the beginning of a stretch of uncharged and hydrophobic amino acids of approximately 25 residues, thought to be a membrane-binding subregion. (Twenty-five residues in the conformation of an α -helix could span 35 Å, the approximate thickness of the hydrocarbon core of a lipid bilayer⁴¹.) At the carboxy-terminal end of the membrane-binding region there is a cluster of basic amino acids (positions 308 and 310-312) that may serve to anchor the H-2 molecule in the membrane through interaction with the negatively charged phosphate groups of the membrane phospholipids. Glycophorin⁴², membrane-bound IgM⁴³ and HLA⁴⁴ have similar clusters of basic residues at the end of the membrane-binding portion of the molecules. That portion of the molecule beginning at the inner face of the membrane and stretching to the C-terminus is apparently located inside the cell⁴⁵ and possible functions in the communication of signals to the cell interior following interactions of the exposed extracellular region of the molecule. Whether the phosphoserine residues purported to be present in this portion of H-2⁴⁶ and HLA⁴⁷ molecules is involved in such processes remains to be demonstrated.

As depicted in the schematic model (Fig. 2), the molecule divides neatly into three functional segments: the extracellular portion which includes the noncovalently bound β_2 M and contains the alloantigenic determinants, the membrane segment that penetrates the lipid bilayer and allows the molecule to span the plasma membrane, and the intracellular segment that possibly serves in the communication of signals to the interior of

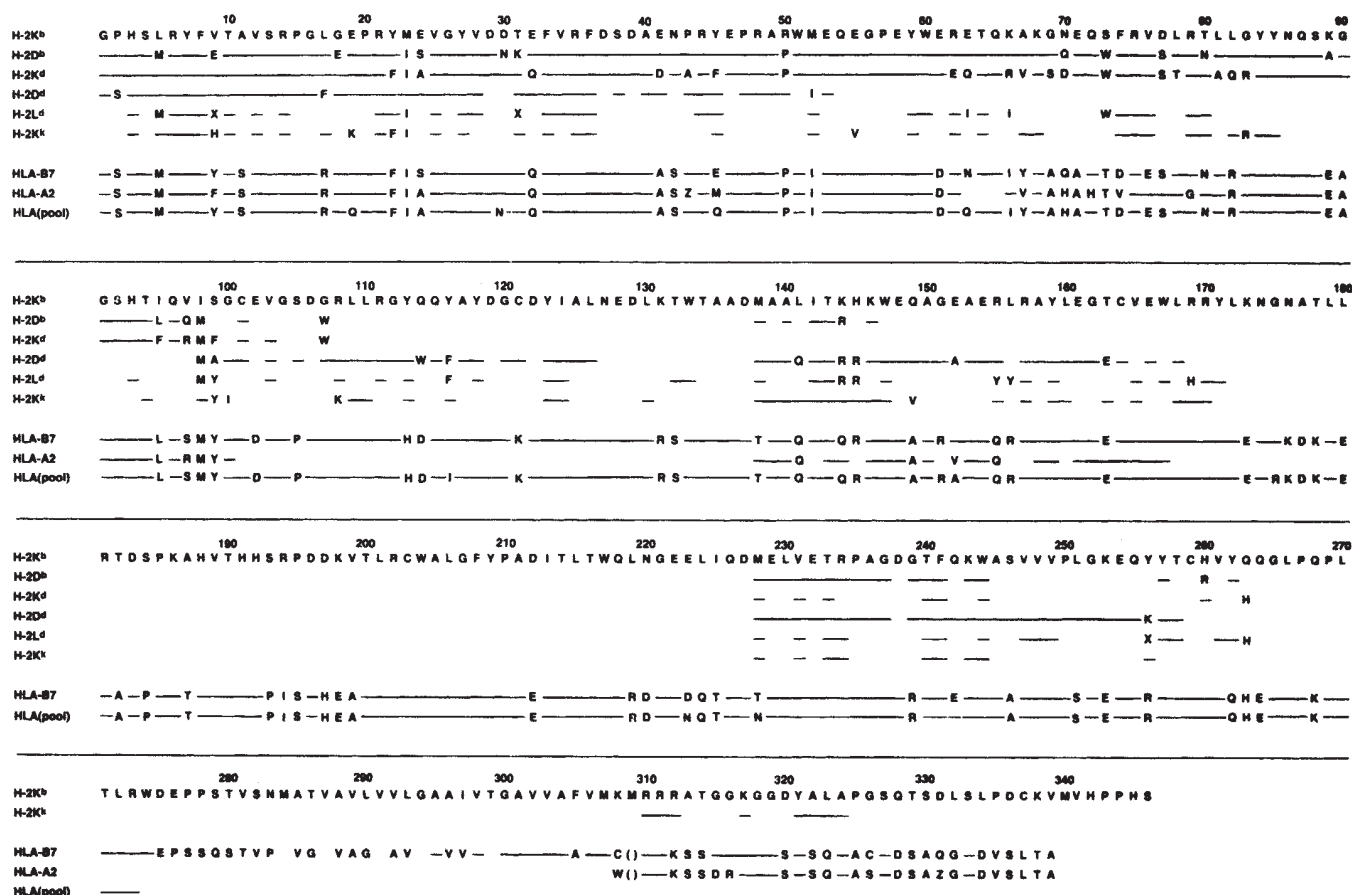


Fig. 3 Compilation of amino acid data for histocompatibility antigens for which extensive data are available. The H-2K^b data have been reported in refs 13, 56; H-2D^b, ref. 17; H-2K^d, ref. 16; H-2D^d, ref. 14; H-2L^d, ref. 15; H-2K^k, ref. 46; HLA-B7 and HLA-A2, refs 10, 59, 60; and pooled HLA, ref. 49. Sequences are given in the single letter amino acid code described in Fig. 1 legend. × indicates an unidentified nonidentical residue. Straight lines indicate identity to the K^b sequence. The parentheses denote a position in the HLA molecules where a deletion must be made in order to achieve maximum alignment to H-2K^b.

	10					20					30					40					50										
Common to H-2 and HLA	G S H S M R Y F * T * V S R P G * G E P R F I A V G Y V D D T Q F V R F D S D A * * P R E E P R A P																														
H-2 variable residues	X		X		X'							X X X		X X X		X X X					X		X		X		X				
HLA variable residues						X							X		X		X										X'		X'		
	60					70					80					90					100										
Common to H-2 and HLA	W I E Q E G P E Y W * R Q T Q I V K * Q * Q * * R V S L R T L * G Y Y N Q S * * G S H T L Q R M Y G																														
H-2 variable residues	X		X		X' X'					X' X		X X'		X		X' X'		X X' X X		X'					X					X' X X X	
HLA variable residues						X'		X'					X		X		X X												X		

Fig. 4 Comparison of the H-2 and HLA amino acid sequences for the NH₂ terminal 100 residues. The residues indicated by single letter code (see Fig. 1 legend) are amino acids shared by at least one H-2 and one HLA molecule. At positions where more than one amino acid residue is common to H-2 and HLA, the predominant residue is listed. Asterisks indicate positions where no residues common to H-2 and HLA molecules have been found, but where shared residues could be created by single base changes in the DNA coding sequences. Boxed asterisks are those positions which require a double base change for the generation of shared residues. × indicates positions in H-2 or HLA where multiple substitutions occur but which only require single base changes in the DNA coding sequences for the generation of observed substitutions: ×' indicates those positions which require multiple base changes for the generation of observed differences.

the cell. Division of the extracellular portion of the molecule into subregions analogous to immunoglobulin-like domains is prompted by primary structural homologies of HLA and β_2 M to IgG constant region domains^{48,49}. The division (Fig. 2) is supported by the location of the two points where the molecule is susceptible to cleavage by the proteolytic enzyme, papain, which suggest an interruption of a domain-like structure of a globular protein. The subregions (domains) are designated N for the NH₂-terminal portion, C1 and C2 for Cys-containing segments.

Relationship of H-2 and immunoglobulin

The model of the H-2K^b molecule showing domain-like subregions suggests a certain kinship to immunoglobulins. This relationship is supported by weak but significant amino acid sequence homologies between HLA, β_2 M and IgG constant regions⁴⁸⁻⁵⁰, and by the facts that the histocompatibility and immunoglobulin molecules both consist of two chains encoded by unlinked genes and that both the 'heavy' chains are products of complex genetic loci. However, there are some very basic differences. For example, there is no evidence for hypervariable regions or extensive somatic diversification in the H-2K^b molecule as has been found in immunoglobulin variable regions. Whereas pooled immunoglobulin would show heterogeneity in amino acid sequence⁵¹, no evidence of microheterogeneity was found in material isolated from E1-4 tumour cells. Furthermore, when compared by peptide mapping, no differences were observed between H-2K^b isolated from C57BL/6 splenocytes⁵² and that isolated from E1-4 tumour cells. In addition, analyses of K^b molecules isolated from splenocytes of mutant mice have revealed sequences identical to those found for the H-2K^b molecules isolated from E1-4⁵³. Only discrete differences (one or two amino acids) were found between the H-2K^b molecules isolated from splenocytes from 10 mutant strains of mice and that from E1-4, and they undoubtedly related to the mutations which allowed detection of these molecules in biological assays⁵³. Furthermore, examination of the data presented below (Fig. 3) reveals that sequence differences among the H-2 molecules from a given haplotype occur in all parts of the molecule and not just in their NH₂-terminal portions, thus precluding extrapolation of the analogy between H-2 and immunoglobulin molecules to include their having in common structures with clearly delineated variable and constant regions.

Comparison of H-2 and HLA

Although H-2K^b is the only major mouse histocompatibility antigen for which a complete structure is known, extensive areas of several other murine H-2 molecules have been determined and nearly complete primary structures have been reported for

the human HLA-B7 and HLA-A2 molecules as well as for a pool of human HLA antigens (Fig. 3). Structural features apparently conserved by all these molecules include the two intra-chain disulphide loops which span positions 101-164 and 203-259 in H-2K^b (ref. 54), and the carbohydrate moiety which is linked to an Asn residue at position 86 (refs 10, 32). Important for the structural studies, which are based on obtaining peptides by CNBr cleavage, is the highly repetitive occurrence of Met residues at identical positions among the different H-2 molecules³².

Comparison of the amino acid sequence of the NH₂-terminal 100 residues from three H-2, two HLA molecules and the pool of HLA molecules, reveals a high degree of primary structural homology among these molecules. The homology of K^b to D^b, K^b to K^d and K^d to D^b is 85%, 75% and 76%, respectively, and the homology between the HLA molecules is 85%. Comparison between H-2 and HLA molecules yields about 70% homology in all cases. Further examination of the NH₂-terminal sequence data indicates that 63 residues are common to all H-2 molecules (Fig. 3). (These data include the partial sequence data available for K^k, D^d and L^d.) Of these 63 residues, 54 (or 86%) are also common to HLA. Even more remarkably, at 86 of the 100 positions at least one H-2 and one HLA molecule share a common amino acid (Fig. 4). In addition, at only 1 of the 14 positions where common amino acids are not shared would the substitutions observed require a multiple base change in the DNA-coding sequences for common amino acids to occur.

Examination of the NH₂-terminal 100 residues of the H-2 molecules reveals that there are 37 positions where at least two different amino acid residues occur in the sequence; most of these differences are nonrandomly distributed. Small clusters of differences are evident at positions 18-24, 30-32, 41-45 and 95-100, and a large concentration of differences is found between positions 61 and 83, these clusters being interspersed with stretches of conserved amino acid residues (Fig. 4). All these regions represent areas of diversity between the murine and human molecules; however, only the large cluster of differences (61-83) represents an area of diversity when only the human molecules are compared. It is tempting to speculate that the amino acid interchanges observed at positions 61-83 represent an important region of diversity among these alloantigens. In support of this possibility, the homology in this region is strikingly lower than the overall NH₂-terminal homology: only four of the 23 positions (17%) are common to the five molecules for which extensive NH₂-terminal sequence data are available, whereas 50 of the remaining 77 residues (65%) are identical. Likewise, only 10 residues (43%) are shared among the three H-2 molecules in region 61-83 whereas 53 (69%) are shared for the remaining 77 residues. Further-

more, 7 of the 10 amino acid conversions which require two base changes at the nucleic acid level for interconversion occur in this region (Fig. 4).

The extensive sequence data available for H-2K^b and HLA-B7 allow comparisons to be made over the entire length of these molecules. Thus, although the overall homology between these molecules is 71% for positions 1–281 (extracellular portion), the homology for the stretch after position 281 drops to 34%. Comparison of K^b and HLA-A2 in positions available for comparison in this region shows 28% homology. In contrast, the homology between the human molecules (87%)¹⁰ is similar to that observed for the extracellular region of the molecule. In addition to the striking degree of sequence difference between the murine and human molecules, the K^b molecule is 10 residues longer than the human molecules. To achieve maximum alignment of the K^b and B7 molecules, a deletion must be placed at position 307 of the B7 molecule. Also, the second carbohydrate chain present at position 176 in K^b is not detected in HLA-B7.

The finding that there is significantly less homology between K^b and K^d (presumed alleles) than between K^b and D^b is intriguing; it agrees with previous conclusions from peptide mapping data³⁰ and is difficult to reconcile with the concept that

the multiple histocompatibility gene loci arose by gene duplication in a common ancestor. In such a model, divergence of the allelic genes would arise through mutations occurring after speciation. The present data suggest that the gene duplication event occurred subsequent to speciation.

Finally, a major goal of the primary structural studies on H-2 molecules is the localization of sites involved in recognition by alloantibodies and by activated T lymphocytes. It is tempting to speculate that the regions of diversity, especially that found between residues 61 and 83, are involved in such processes. However, although the differences outlined above between the H-2 molecules are quite dramatic, recent data on the structural characterization of H-2 mutants suggest that only small numbers of amino acid differences between antigens are required for immune recognition. Differences between parental and mutant molecules at one or two positions are sufficient to elicit a highly responsive and specific immune recognition, both in alloaggression and in 'associative recognition'^{53,55}. Therefore, the multiple differences among the H-2 molecules shown above provide a potentially vast repertoire of recognition determinants for both T-cell allorecognition and serological distinctions.

Parts of this study were supported by NIH grants AI 07289 and AI 10702.

Received 12 January; accepted 18 February 1981.

- Paul, W. E. & Benacerraf, B. *Science* **195**, 1293–1299 (1977).
- Klein, J. *Science* **203**, 516–521 (1979).
- Klein, J. *Biology of the Mouse Histocompatibility-2 Complex* (Springer, New York, 1975).
- Snell, G. D., Dausset, J. & Nathenson, S. G. *Histocompatibility* (Academic, New York, 1976).
- Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27**, 51–177 (1979).
- Shearer, G. M. & Schmidt-Verhulst, A. M. *Adv. Immun.* **25**, 55–91 (1977).
- Klein, J. *Adv. Immun.* **26**, 55–146 (1978).
- Festenstein, H. & Démar, P. *Curr. Topics Immun.* **9** (1979).
- Nathenson, S. G., Uehara, H., Ewenstein, B., Kindt, T. J. & Coligan, J. E. *Rev. Biochem.* (in the press).
- Strominger, J. L. *et al.* in *The Role of the Major Histocompatibility Complex in Immunobiology* (eds Benacerraf, B. & Cantor, H.) (Garland, New York, 1981).
- Vitetta, E. S. & Capra, J. D. *Adv. Immun.* **26**, 147–193 (1978).
- Uehara, H. *et al. Biochemistry* **19**, 6182–6188 (1980).
- Martinko, J. M. *et al. Biochemistry* **19**, 6188–6193 (1980).
- Nairn, R., Nathenson, S. G. & Coligan, J. E. *Eur. J. Immun.* **10**, 495–503 (1980).
- Coligan, J. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 1134–1138 (1980).
- Kimball, E. S., Nathenson, S. G. & Coligan, J. E. *Biochemistry* (in the press).
- Maloy, W. L., Nathenson, S. G. & Coligan, J. E. *J. biol. Chem.* **256**, 2863–2872 (1981).
- Ewenstein, B. M., Freed, J. H., Mole, L. E. & Nathenson, S. G. *Proc. natn. Acad. Sci. U.S.A.* **73**, 915–918 (1976).
- Ewenstein, B. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 2909–2913 (1978).
- Coligan, J. E. *et al. Molec. Immun.* **16**, 3–8 (1979).
- Uehara, H. *et al. Biochemistry* **19**, 306–315 (1980).
- Gates, F. T. III, Coligan, J. E. & Kindt, T. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 554–558 (1981).
- Kindt, T. J., Coligan, J. E., Uehara, H. & Nathenson, S. G. *Proc. int. Conf. Chemical Synthesis of Peptides and Proteins* (Elsevier, Amsterdam, in the press).
- Schwartz, B. D., Kato, K., Cullen, S. E. & Nathenson, S. G. *Biochemistry* **12**, 2157–2164 (1973).
- Henrikson, O., Robinson, E. A. & Appella, E. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3322–3326 (1978).
- Rogers, M. J., Robinson, E. A. & Appella, E. *J. biol. Chem.* **254**, 1126–1133 (1979).
- Freed, J. H., Sears, D. L., Brown, J. L. & Nathenson, S. G. *Molec. Immun.* **16**, 9–21 (1979).
- Silver, J. & Hood, L. *Nature* **256**, 63–64 (1975).
- Capra, J. D., Vitetta, E. S., Klapper, D. G., Uhr, J. W. & Klein, J. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3661–3664 (1976).
- Brown, J. L., Kato, K., Silver, J. & Nathenson, S. G. *Biochemistry* **13**, 3174–3178 (1974).
- Maloy, W. L., Hämmerling, G., Nathenson, S. G. & Coligan, J. E. *J. immun. Meth.* **37**, 287–299 (1980).
- Kimball, E. S., Maloy, W. L., Martinko, J. M., Nathenson, S. G. & Coligan, J. E. *Molec. Immun.* **17**, 1283–1291 (1980).
- Coligan, J. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3390–3394 (1978).
- Kimball, E. S., Coligan, J. E. & Kindt, T. J. *Immunogenetics* **8**, 201–211 (1979).
- Gates, F. T. III, Kimball, E. S. & Coligan, J. E. (in preparation).
- Michaelson, J., Rothenberg, E. & Boyse, E. A. *Immunogenetics* **11**, 93–95 (1980).
- Goding, J. W. & Walker, I. D. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7395–7399 (1980).
- Robinson, P. J., Graf, L. & Sege, K. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Nathenson, S. G. & Shimada, A. *Transplantation* **6**, 662–663 (1968).
- Mann, D. L., Rogentine, G. N. Jr, Fahey, J. L. & Nathenson, S. G. *Nature* **217**, 1180–1181 (1968).
- Dickerson, R. E. & Geis, I. *The Structure and Action of Proteins*, 28 (Harper & Row, London, 1969).
- Tomita, M., Furthmayr, H. & Marchesi, V. T. *Biochemistry* **17**, 4756–4770 (1978).
- Rogers, J. *et al. Cell* **20**, 303–312 (1980).
- Robb, R. J., Terhorst, C. & Strominger, J. L. *J. biol. Chem.* **253**, 5319–5324 (1978).
- Walsh, F. S. & Crumpton, M. J. *Nature* **269**, 307–311 (1977).
- Rothbard, J. B., Hopp, T. P., Edelman, G. M. & Cunningham, B. A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4239–4243 (1980).
- Pober, J. S., Guild, B. C. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 6002–6004 (1978).
- Orr, H. T., Lancet, D., Robb, R. J., Lopez de Castro, J. A. & Strominger, J. L. *Nature* **282**, 266–270 (1979).
- Trägårdh, L., Rask, L., Wiman, K., Fohlman, J. & Peterson, P. A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1129–1133 (1980).
- Coligan, J. E. *et al. Transplant Proc.* (in the press).
- Kabat, E. A., Wu, T. T. & Bilofsky, H. *Sequences of Immunoglobulin Chains* (NIH Publ. No. 80-2008, 1979).
- Nisizawa, T., Ewenstein, B. M., Uehara, H., McGovern, D. & Nathenson, S. G. *Immunogenetics* **12**, 33–44 (1981).
- Nairn, R., Yamaga, K. & Nathenson, S. G. *Rev. Genet.* **14**, 241–277 (1980).
- Martinko, J. M., Halpern, R., Aldersberg, J. R. & Nathenson, S. G. *Molec. Immun.* (in the press).
- Ewenstein, B. M. *et al. Immunogenetics* **11**, 383–396 (1980).
- Uehara, H., Coligan, J. E. & Nathenson, S. G. *Biochemistry* (submitted).
- Nathenson, S. G. *et al.* in *Current Trends in Histocompatibility* (eds Ferrone, S. & Reisfeld, R. A.) (Plenum, New York, in the press).
- Trägårdh, L., Wiman, K., Rask, L. & Peterson, P. A. *Biochemistry* **18**, 1322–1328 (1979).
- Orr, H. T., Lopez de Castro, J. A., Parham, P., Ploegh, H. L. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4395–4399 (1979).
- Orr, H. T., Lopez de Castro, J. A., Lancet, D. & Strominger, J. L. *Biochemistry* **18**, 5711–5720 (1979).

Major rearrangement in the human β -globin gene cluster

R. W. Jones, J. M. Old, R. J. Trent, J. B. Clegg & D. J. Weatherall

MRC Molecular Haematology Unit, Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK

A form of $\delta\beta$ -thalassaemia results from a major chromosomal rearrangement involving the β -globin gene cluster which includes an inversion of the entire sequence between the γ - and δ -globin genes and two deletions.

IN MAN the ϵ , γ , γ , δ - and β -globin genes that code for the embryonic, fetal and adult β -like globin chains of haemoglobin are closely linked on chromosome 11^{1–6}. The hereditary disorders which result in their altered expression include forms of

pancellular hereditary persistence of fetal haemoglobin (HPFH) and $\delta\beta$ -thalassaemia which are characterized by an absence of δ - and β -chain production and a variable persistence of γ , γ , or γ - and γ -chain synthesis into adult life.

We describe here the analysis of the ϵ - γ - δ - β gene cluster in DNA of an Indian patient (FB) who has previously been shown by haematological, haemoglobin and family studies to be homozygous for ϵ - γ - δ -thalassaemia⁷. Restriction endonuclease analysis indicates that the disorder results from a

Table 1 Size of restriction enzyme fragments

Enzyme	Size in normal DNA (kb)	Size in ϵ - γ - δ -thalassaemia DNA (kb)
a Detected with γ-cDNA probe		
<i>Bam</i> HI	2.6+5.1+15.5	2.6+5.1
<i>Bgl</i> II	13	10.5
<i>Eco</i> RI	7.2+1.6+2.7	7.2+1.6+4.2
<i>Hpa</i> I	24+5.1+14	24+4.4
<i>Pst</i> I	4.0+5.1	4.0+5.1
<i>Xho</i> I	5.1	5.1
b Detected with β-cDNA probe		
<i>Bam</i> HI	15.5+4.4+1.8+8.1	20
<i>Bgl</i> II	7.7+5.0	10.5+6.6
<i>Eco</i> RI	2.2+1.7+5.4+3.6	4.2+3.6
<i>Hpa</i> I	1.9+1.4+7.0	1.9+17

DNA was isolated from peripheral blood of a patient (FB) homozygous for ϵ - γ - δ -thalassaemia and from a normal human spleen by phenol chloroform extraction²⁹. DNA samples were digested with restriction endonucleases as indicated and fractionated in a vertical 0.8% agarose gel by electrophoresis for 15 h at 1.5 V cm⁻¹. The DNA fragments were transferred to a nitrocellulose filter by the gel blotting technique of Southern³⁰ and the nitrocellulose filter hybridized³¹ to a cloned γ - or β -cDNA probe labelled with ³²P by nick translation³². After hybridization at 42 °C for 48 h the filter was washed in 0.1 × SSC at 52 °C as described elsewhere³³ and the bands detected by autoradiography³⁴. The γ -cDNA probe was the γ -globin cDNA plasmid JW151 (ref. 35); the β -cDNA probe was β -globin cDNA plasmid JW102 (ref. 35). The 3.6-kilobase *Eco*RI fragment hybridizing with the β -cDNA probe did not have a place in the final structure of FB DNA. This was probably due to hybridization of the β -cDNA probe to the ϵ gene¹.

previously undescribed chromosomal rearrangement involving two deletions totalling 8.5 kilobases of DNA and an inversion of 15.5 kilobases of DNA in the γ - δ - β gene cluster. Unlike the deletions previously described in δ -thalassaemia^{4,8-10} or HPFH^{4,11,12}, the entire sequence between the γ - and δ -globin genes is retained but is inverted. This is the first example of a major inversion as the basis for a human hereditary disease.

An abnormal γ gene, and abnormal δ - or β -globin sequences

Hybridization of several single (Table 1a) and double (not shown) restriction enzyme digests of FB DNA to a cloned γ -cDNA probe showed a normal arrangement of restriction sites on the 5' side of the γ gene up to the *Pst*I site in that gene (A in Fig. 1). Restriction enzymes producing DNA fragments which span this point in normal DNA produced abnormal fragments when used to digest FB DNA, which suggested that DNA on the 3' side of A (Fig. 1) had been deleted. However, hybridization of a cloned β -cDNA probe to single enzyme digests of FB DNA (Table 1b) showed a number of DNA fragments, most of which differed in size from those seen in normal DNA. Among these were a 4.2-kilobase *Eco*RI and a 10.5-kilobase *Bgl*II fragment, both of which were the same size as abnormal fragments found in digests with the same enzymes hybridized to the γ -cDNA probe (Table 1a). This suggested that there was one abnormal DNA fragment generated by each enzyme which spanned the remaining 5' half of the γ gene and also included some part of either the δ or β genes.

To test this hypothesis, the 4.2-kilobase *Eco*RI and 10.5-kilobase *Bgl*II fragments were digested with *Xho*I. The latter cuts once in the γ gene and once in the γ gene very close to A (Fig. 1) but nowhere else in the whole ϵ - γ - δ - β region. Thus the

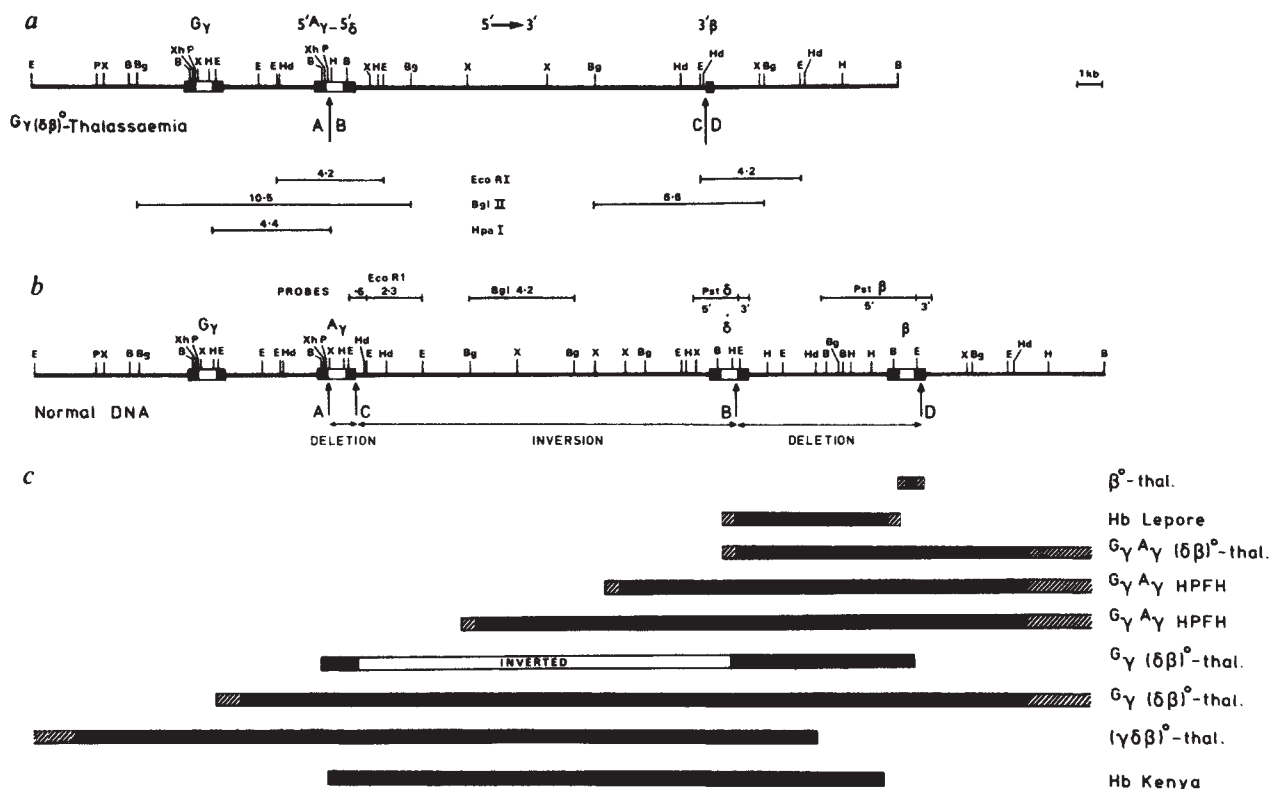


Fig. 1 A map of restriction endonuclease sites in and around the β -like globin gene region of ϵ - γ - δ -thalassaemia DNA compared with normal DNA. **a**, Restriction map of DNA from a patient with ϵ - γ - δ -thalassaemia. The brackets below the map indicate the locations of some of the abnormal fragments listed in Table 1. **b**, Restriction map of normal DNA derived from refs 4, 24 and 25, except for the *Hind*III site on the 5' side of the δ -globin gene, which was located by hybridization of the 2.3-kilobase *Eco*RI fragment to FB DNA (Fig. 5) and to normal DNA (data not shown). A, B, C and D mark the boundaries of the normal sequences which have been deleted and inverted in the ϵ - γ - δ -thalassaemia DNA. The brackets above the normal DNA map indicate the positions of the *Eco*RI 0.6, *Eco*RI 2.3, *Bgl*II 4.2, *Pst* δ and *Pst* β fragments. **c**, Schematic map of the deletions in the β -like globin region in various diseases. The deletions are shown by solid boxes and the end points are located within the hatched boxes where known. The maps are from the following refs: β -thalassaemia, ref. 26; Hb Lepore, 2,27; ϵ - γ -HPFH, 4,11,12; ϵ - γ - δ -thalassaemia, 4,8,9; ϵ - γ - δ -thalassaemia, 4,10; (γ - δ)-thalassaemia, 21. Different locations of the 3' end of the ϵ - γ - δ -thalassaemia deletion have been proposed^{4,8,9} and its precise location is unknown. Hb Kenya has not been proved to be a deletion by restriction enzyme mapping. Restriction enzyme sites are: B, *Bam*HI; BG, *Bgl*II; E, *Eco*RI; Hd, *Hind*III; H, *Hpa*I; P, *Pst*I; X, *Xba*I; and Xh, *Xho*I.

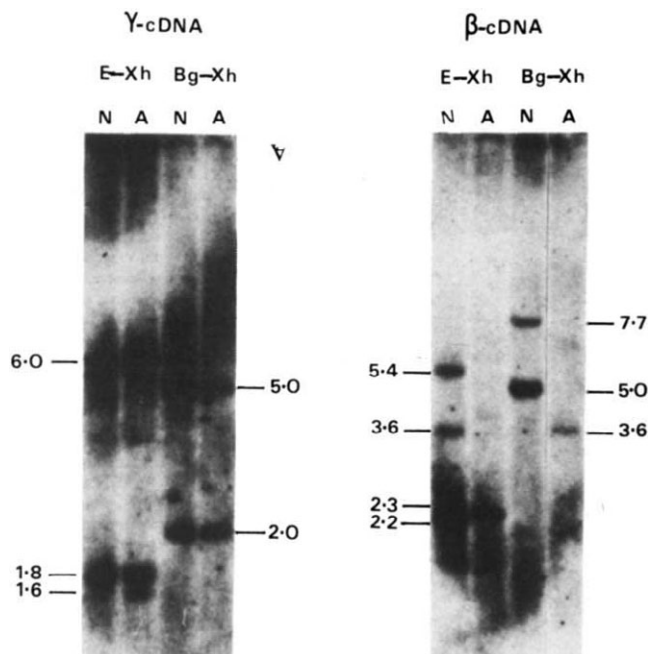


Fig. 2 Hybridization of double digests of normal and FB DNA to cloned γ - and β -cDNA probes. Normal (N) and FB DNAs (A) were digested with either *Xho*I and *Eco*RI (E-Xh) or with *Bgl*II and *Xho*I (Bg-Xh), and duplicate filters were hybridized to cloned γ - and β -cDNAs as described in Table 1 legend. Although a proportion of *Xho*I sites in DNA from some tissues are modified by methylation and therefore are resistant to digestion, there was no evidence of this in the DNA digested here. The size in kilobases of each fragment is shown.

4.2-kilobase *Eco*RI fragment should be cleaved by *Xho*I to give one γ gene fragment and one β -like gene fragment. Double digests of FB DNA with *Eco*RI and *Xho*I (Fig. 2) generated a 1.8-kilobase fragment which hybridized only to the γ -cDNA probe and a 2.3-kilobase fragment which hybridized only to the β -cDNA probe. The sum of these two fragments is 4.1 kilobases, which is close to the expected sum of 4.2. Similarly, *Xho*I cut the *Bgl*II 10.5-kilobase fragment twice to give fragments of 5.0 and 2.0 kilobases which hybridized to γ -cDNA, but only one fragment of 3.6 kilobases on hybridization with β -cDNA (Fig. 2). The sum of these three fragments is 10.9 kilobases, again in close agreement to the predicted value of 10.5.

Both these double digests can therefore be interpreted as showing that in FB DNA, the normal γ -gene sequence changes to one containing either all or part of either the β or δ gene, or parts of both. Some β -like sequence must lie within 2.3 kilobases of the *Xho*I site in the γ -gene to account for the presence of a DNA fragment of that size in *Eco*RI/*Xho*I digests hybridized to cloned β -cDNA probe.

Inversion of DNA sequence

The relative positions of several new restriction sites, including those of *Bgl*II and *Eco*RI, appeared to be the reverse of those on the 5' side of the δ -globin gene in normal DNA. An inversion of the normal DNA sequence between A and B (Fig. 1b) would place the 5' half of the δ gene next to the 5' half of the γ gene, and generate 3.9-kilobase *Eco*RI and 10.7-kilobase *Bgl*II fragments which would hybridize to both γ - and β -cDNA probes (Fig. 1a). To determine which β -like sequence lay close to A, 5'- and 3'- δ and β probes were hybridized to *Eco*RI and *Bgl*II digests (Fig. 3). Neither the 3'- δ nor 5'- β probes hybridized to FB DNA, suggesting that these DNA sequences had been lost in the rearrangement. The 5'- δ probe hybridized to 4.2-kilobase *Eco*RI and 10.5-kilobase *Bgl*II fragments supporting the presence of the 5'- δ sequence close to the 5' half of the γ gene. However, the 3'- β probe hybridized to a 4.2-kilobase *Eco*RI fragment as well. As it also hybridized to a

6.6-kilobase *Bgl*II fragment, rather than one of 10.5 kilobases, it seemed likely that the 3'- β sequence was involved in some further rearrangement of the normal DNA sequence such as the point of reattachment of an inverted sequence to the chromosome (junction C/D, Fig. 1a).

The *Eco*RI 2.3- and 0.6-kilobase probes are generated from the normal DNA sequence close to the 3' end of the γ gene (Fig. 1b). Both these probes, as well as the 3'- β probe, hybridized to the 6.6-kilobase *Bgl*II fragment from FB DNA (Fig. 4), showing that this fragment contains sequences complementary to those at both C and D in Fig. 1b. The 4.2-kilobase *Eco*RI fragment that hybridized with the 5'- δ probe contains no *Bgl*II site, distinguishing it from the *Eco*RI fragment of the same size to which both the *Eco*RI 0.6-kilobase and 3'-*Pst* β probes hybridized (Fig. 4).

This confirms the separate identities of the two abnormal 4.2-kilobase *Eco*RI fragments, and demonstrates that the relative positions of *Eco*RI and *Bgl*II restriction sites around the junction C/D (Fig. 1a) are consistent with the proposed DNA rearrangement.

Linkage of γ , δ and β sequences

To establish the overall arrangement of the sequences in FB DNA, single and double digests with *Hind*III and *Xba*I were hybridized to a series of probes complementary to DNA sequences spanning the γ and δ genes (Fig. 5). Hybridization of these digests to the γ -cDNA probe supported the normal location of *Hind*III and *Xba*I restriction sites on the 5' side of the γ gene. The same digests also demonstrated the presence of abnormal 7.2-kilobase *Xba*I and 18.5-kilobase *Hind*III fragments which must include the 5' half of the γ gene, and extend

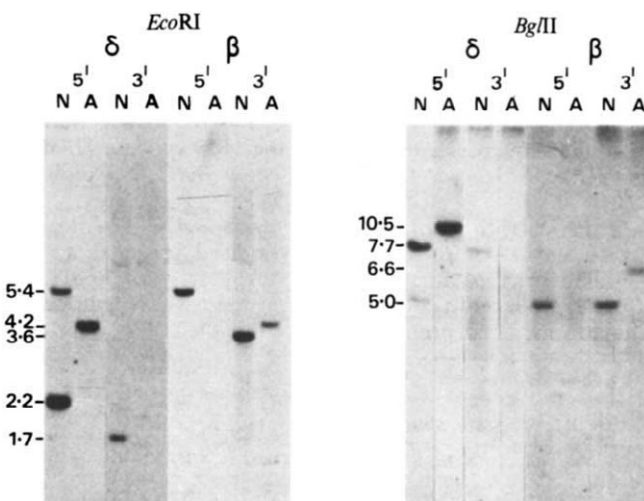


Fig. 3 Hybridization of *Eco*RI and *Bgl*II digests of normal and γ -(δ)-thalassaemia DNA to 5'- and 3'-*Pst* δ and 5'- and 3'-*Pst* β probes. Normal and FB DNAs were digested with either *Eco*RI or *Bgl*II and duplicate filters hybridized with each of these probes as described in Table 1 except that the hybridization buffer contained 10% dextran sulphate; hybridization was for 12 h, and the filters were finally washed in $0.1 \times$ SSC at 65 °C (ref. 28). The probes were prepared by agarose gel electrophoresis of *Eco*RI digests of the cloned *Pst* β or *Pst* δ fragments. The DNA fragments containing the sequence on either the 5' or 3' side of the *Eco*RI site within the δ -globin gene and the β -globin gene were identified after staining with ethidium bromide and eluted electrophoretically from the agarose gel. The size in kilobases of each fragment is shown. *Eco*RI cuts inside the normal δ -globin gene to give a 2.2-kilobase fragment containing 5' sequences and a 1.7-kilobase fragment containing 3' sequences; inside the normal β -globin gene it gives a 5.4-kilobase fragment containing 5' sequences and a 3.6-kilobase fragment containing 3' sequences. *Bgl*II cuts outside the normal δ - and β -globin genes to give a 7.7-kilobase fragment containing both 5' and 3' δ -globin sequences and a 5.0-kilobase fragment containing both 5' and 3' β -globin sequences. In *Eco*RI digests of normal DNA there is no significant cross-hybridization of the 5' probes to 3' sequences, or vice versa. There is limited cross-hybridization between two 5' probes, and between the two 3' probes, but the absence of any significant hybridization of the 5'-*Pst* β or 3'-*Pst* δ probes to FB DNA eases any problems of interpretation. N, normal DNA; A, FB DNA; δ , *Pst* δ probe; β , *Pst* β probe.

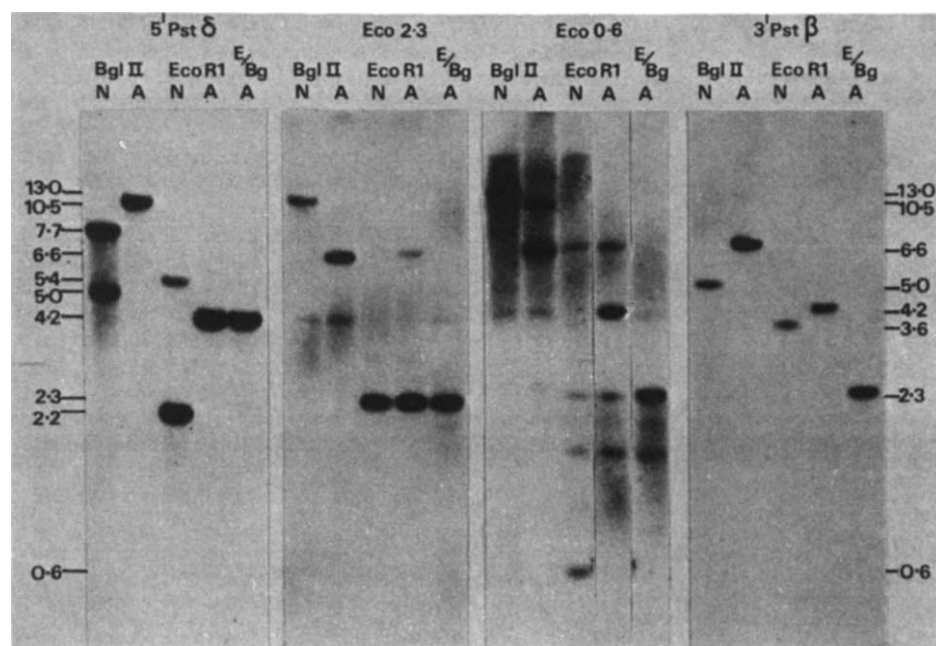


Fig. 4 Hybridization of FB DNA with 5'-Pst δ , EcoRI 2.3, EcoRI 0.6 and 3'-Pst β probes. The 3'- and 5'-Pst probes were prepared as described in Fig. 3, and the EcoRI 0.6 and 2.3 probes in the same way after digestion of λ H γ G4⁵ with EcoRI. The normal (N) and FB (A) DNAs were digested with EcoRI, BglII, or with both enzymes (E/Bg), and duplicate filters were hybridized to each of these probes as described in Fig. 3 legend. The sizes of DNA fragments in kilobases are indicated. The EcoRI 0.6-kilobase probe hybridizes to part of the 3' coding sequence of both γ genes. Thus in EcoRI digests of normal and FB DNA, it hybridizes to a 1.6-kilobase fragment containing the 3' end of the γ gene as well as to a 0.6-kilobase fragment containing the 3' end of the δ gene (normal DNA) and a 4.2-kilobase fragment (FB DNA). In BglII digests, as well as hybridizing to a 6.6-kilobase fragment in FB DNA, this probe hybridized to the abnormal 0.5-kilobase fragment containing the intact γ gene. In addition, cross-hybridization occurred with a number of unidentified fragments present in both normal and FB DNA.

downstream from it. Both of these fragments were seen when these digests were hybridized with the 5'-Pst δ probe, but only the 18.5-kilobase HindIII fragment also hybridized with the 4.2-kilobase BglII and 2.3-kilobase EcoRI probes (Fig. 5). The sequences hybridizing to these two probes must therefore lie downstream from the δ and γ sequences. Because the 2.3 EcoRI probe hybridized with a 0.9-kilobase HindIII fragment, as well as with the 18.5-kilobase HindIII fragment, the sequence hybridizing with this probe must straddle the HindIII site at the 3' end of the large HindIII fragment. As we already know the relative orientation of the 5'- δ and 5'- γ sequences, we can order the sequences hybridizing to the 18.5-kilobase HindIII fragment as: 5'- δ ; 5'-Pst δ ; 4.2-kilobase BglII; 2.3-kilobase EcoRI.

In digests with XbaI, the 4.2-kilobase BglII, 2.3-kilobase EcoRI and 3'-Pst β probes all hybridized to an abnormal 9.0-kilobase XbaI fragment. Double digests with XbaI and HindIII produced a 5.6-kilobase fragment which hybridized with the first two probes, but not with the 3'-Pst β probe.

Therefore the 9-kilobase XbaI fragment overlaps the 3' end of the 18.5-kilobase HindIII fragment by 5.6 kilobases, and the 3'-Pst β probe must hybridize to a sequence downstream from this overlap, on the 3' side of the sequence hybridizing with the 2.3-kilobase EcoRI probe.

More precise information about the relative positions of the sequences hybridizing to the 4.2-kilobase BglII, 2.3-kilobase EcoRI and 3'-Pst β probes was provided by these single and double digests. The BglII probe hybridized to a 2.0-kilobase XbaI fragment as well as to the 9.0-kilobase XbaI fragment, showing that this sequence straddled the 5' end of the latter fragment. The sequence hybridizing to the 2.3-kilobase EcoRI probe must lie near the middle of the 9.0-kilobase XbaI fragment straddling one of the two HindIII sites. The 3'-Pst β probe hybridized with a 4.7-kilobase HindIII fragment which, as shown by a HindIII-XbaI double digest, overlaps the 3' end of the 9.0-kilobase XbaI fragment by 2.4 kilobases, with all the sequence hybridizing with this probe contained in this last fragment.

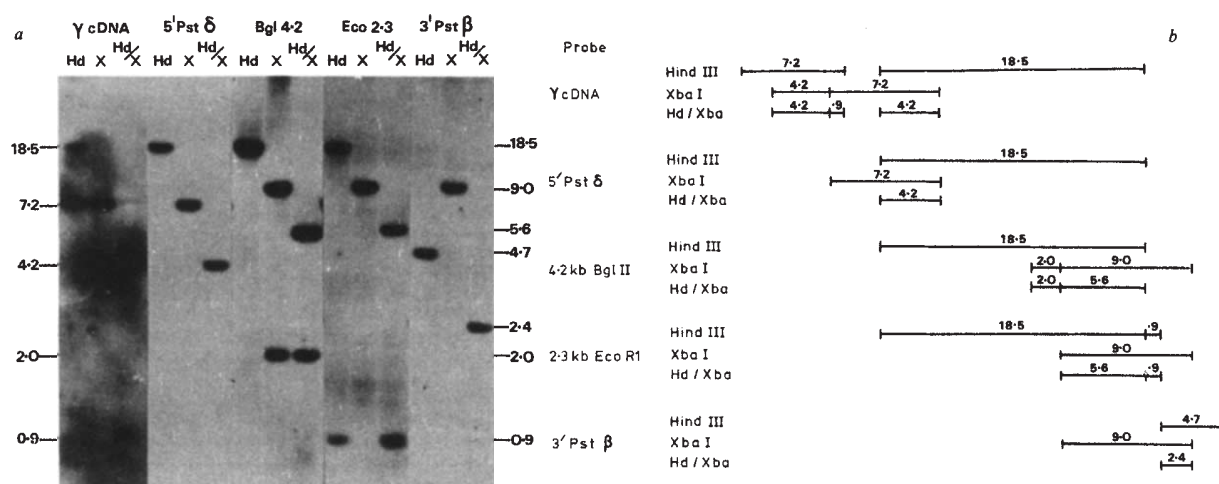


Fig. 5 a, Hybridization of FB DNA with the cloned γ -cDNA, 5'-Pst δ , 4.2 BglII, 2.3 EcoRI and 3'-Pst β probes. The 5'-Pst δ , 3'-Pst β and 2.3 EcoRI probes were prepared as described in the legends to Figs 3 and 4, and the 4.2-kilobase BglII fragment was prepared in the same way after digestion of H γ G4 (ref. 5) with BglII. FB DNA was digested with HindIII (Hd), XbaI (X), or with both enzymes (Hd-X) and duplicate filters hybridized to each of these probes as described in Fig. 3 legend. b, The interpretation of the hybridization data is shown as a series of overlapping fragments spanning the region between the 5' end of the δ gene and the 3' β sequence (see Fig. 1). The size of each fragment is shown in kilobases (kb).

The order of these sequences in FB DNA is therefore 5'- γ ; 5'-*Pst* δ ; 4.2-kilobase *Bgl*II; 2.3-kilobase *Eco*RI; 3'-*Pst* β . In this context it is unlikely that the 0.6-kilobase *Eco*RI probe (Fig. 4) which is contiguous with the 2.3-kilobase *Eco*RI probe in normal DNA, lies anywhere else but between the 2.3-kilobase *Eco*RI and 3'- β sequences. The relative position of these sequences, and of the *Xba*I and *Hind*III sites, in this region of FB DNA is entirely consistent with the proposed structure (Fig. 1a).

The completed map allows the points A, B, C and D (Fig. 1) to be localized quite precisely; A lies between *Pst*I and *Xba*I sites near the 5' end of the major intervening sequence in the γ gene, and B is between *Hpa*I and *Eco*RI sites at the 3' end of the δ major intervening sequence. C lies in a 0.6-kilobase *Eco*RI fragment at the 3' end of the γ gene, and, as a cloned β -cDNA probe detected abnormal fragments containing the 3' end of the β gene, D must lie between an *Eco*RI site in the 3' coding sequence of the β gene and the end of the gene.

Mechanism of rearrangement of the non- α gene cluster

The form of $\delta\beta$ -thalassaemia described here seems to result from a major rearrangement of the non- α globin gene cluster which has involved the inversion of 15.5 kilobases of DNA between the γ and δ genes and deletions of approximately 1 kilobase from the γ gene and 7.5 kilobases from between the δ and β genes. This is quite unlike previous findings for $\delta\beta$ -thalassaemia^{4,8-10} and HPFH^{4,8,11,12} which have been shown to result from simple deletions of varying sizes. The rearrangement found in the present case must have resulted from a more complicated mechanism.

It has been suggested that inversions are created between inverted repeat sequences¹³⁻¹⁵. The present arrangement must have involved at least one inversion, but computer analysis of the known sequences close to points A, B, C and D¹⁶⁻¹⁸ (Fig. 1) does not reveal any extensive inverted repeat sequences. Nevertheless, the most economical models for the generation of this rearrangement exclusively involve recombination events of the same kind as would occur between one pair of inverted repeats at A and B, and between a second pair at C and D.

In the case of an intrachromosomal rearrangement, a cross-over between A and B (Fig. 1b) in the γ and δ genes of a single chromosome leads to the inversion of the sequence between these points. This inversion brings the sequence between A and C into continuity with that between B and D, and in the process inverts the repeat sequence at C. C and D are now direct repeats and a second cross-over between C and D deletes the sequences between A and C and B and D, producing the abnormal sequence found here. Alternatively, an inversion between C and D is followed by a deletion between A and B. In this model, both cross-overs are of the same kind and each could have occurred independently giving rise to an inversion within a monocentric chromosome.

An alternative explanation is that the cross-overs were between two chromosomes. The repeat sequences at A and D in the γ and β genes of one chromosome are aligned with those at B and C, respectively, in the δ and γ genes of a second. A cross-over between either A and B or D and C leads to an inversion with the products consisting of one bicentric and one acentric chromosome, preventing the formation of viable daughter cells. A second cross-over between the second pair of inverted repeat sequences is therefore obligatory, generating two monocentric chromosomes, one with the arrangement found here.

Despite their economy, these intra- and inter-chromosomal mechanisms still pose problems. Whatever the probability of a single cross-over, the likelihood of two occurring within 20 kilobases of DNA must be smaller. In the case of two cross-overs between chromosomes the probability is even less because they must both be of the same kind and occur at the same time. One solution is that both cross-overs were closely linked as coincident events in a single chromosome (Fig. 6). In this model both

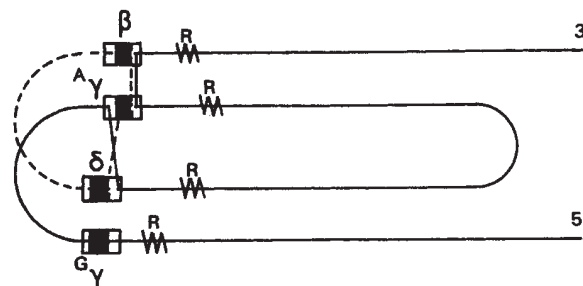


Fig. 6 An intrachromosomal mechanism for the generation of an inversion and deletion. The γ -, δ - and β -globin genes on a single chromosome are aligned so that the cross-overs necessary to generate the inversion and deletion are facilitated. The alignment is brought about by arranging the DNA in a series of loops. The repeat sequences⁵ (R) which lie on either side of the pair of γ genes, and at similar distances on either side of the δ and β genes, might interact and serve to stabilize this looped structure. The distances between the repeats and the nearest gene on the left are to scale. If either cross-over occurs alone, an inversion results. When both occur in conjunction, an inversion and deletion as described in this case of $\gamma(\delta\beta)^{\circ}$ -thalassaemia (continuous line) and an acentric fragment (broken line) are formed.

cross-overs involve a single γ gene, and are therefore separated by no more than 1 kilobase of DNA. This degree of proximity might allow either cross-over to promote the second, or conceivably both may have been part of a single process—this would account for the fact that they are of the same kind.

Regulation of γ - and β -chain synthesis

$\delta\beta$ -Thalassaemia and HPFH constitute a spectrum of conditions in which there is significant γ -chain synthesis in adult life¹⁹; in $\delta\beta$ -thalassaemia this is insufficient to compensate for the absence of δ - and β -chain production, whereas in HPFH γ -chain synthesis is more efficient so that there is less imbalance of globin-chain production.

The deletion which gives rise to $\gamma\gamma^{\circ}\delta\beta$ -thalassaemia leaves the 5' end of the δ -globin gene intact whereas the differently sized deletions that produce $\gamma\gamma^{\circ}\delta\beta$ -HPFH all include the 5' end of the δ -globin gene and variable regions upstream from it (Fig. 1c). It has been suggested, therefore, that there may be a region of DNA upstream from the δ -globin gene which is involved in the regulation of the switch from γ - to β -chain synthesis^{4,8-12,20}. Loss of this region in the various deletions which produce HPFH might result in failure of the neonatal switch from γ - to β -chain production and hence persistent activity of the γ and γ loci at a level approaching that which occurs in fetal life. In the case of $\gamma(\delta\beta)^{\circ}$ -thalassaemia, a single γ -globin gene does not compensate completely for absent δ - and β -chain production, but there is a high level of γ -chain synthesis which must be related to the underlying molecular lesion.

In the one previously reported case of $\gamma(\delta\beta)^{\circ}$ -thalassaemia that has been characterized by restriction enzyme analysis there is a deletion involving all the sequences downstream from a point between the γ and γ genes^{4,10} (Fig. 1c). The phenotypes of this and the present case of $\gamma(\delta\beta)^{\circ}$ -thalassaemia are very similar but of the sequences missing from the first, those between the γ and δ genes, and downstream from the β gene, are intact in our case. If the loss of putative regulatory sequences in either of these regions is responsible for persistent γ -globin synthesis in the first case of $\gamma(\delta\beta)^{\circ}$ -thalassaemia, then the inversion of the sequence between the γ and δ genes, or the interposition of this rearranged sequence between the γ gene and the region downstream from the β gene, must have the same effect in ours. Any regulatory sequences in these regions must therefore be inactivated by the rearrangement of DNA around them, which implies that they are acted on, or act, in *cis*. This comparison does not rule out the existence of regulatory sequences in these regions, but it does place constraints on their nature.

The observation that an intact β -globin gene is inactive in a case of $(\gamma\delta\beta)^{\circ}$ -thalassaemia as the result of a major deletion on

its 5' side suggests that the expression of genes in the ϵ - γ - δ - β cluster is influenced through some form of interaction in *cis*^{11,21}. The sequential arrangement of the ϵ , γ , δ and β genes parallels ontogenesis, and there is recent evidence that γ -globin synthesis precedes δ - and β -globin synthesis in the maturing erythrocyte^{22,23}. This suggests the possibility that these genes are sequentially activated and inactivated by a mechanism, such as a progressive change in chromatin structure, acting only within the chromosome. It would then be expected that at least some of the genetic lesions affecting globin gene activity would be polar in nature, exerting their effect by blocking the progress of gene activation. Thus the common feature in both kinds of ϵ - γ - $(\delta\beta)$ -thalassaemia would be that the deletions start in a similar place, and the precise form of genetic rearrangement downstream from this site would be largely irrelevant. This idea could also be

applied to the non-deletion forms of HPFH in which varying proportions of γ - and β -globin chains are present in the mature adult erythrocyte. Here genetic lesions undetectable by restriction enzyme analysis could, because of their position in the ϵ - γ - δ - β globin cluster, delay the progress of gene activation during erythrocyte maturation so that the mature erythrocyte would contain a higher than normal proportion of γ -globin.

We thank A. B. Amin for providing the ϵ - γ - $(\delta\beta)$ -thalassaemia blood samples, T. Maniatis for the clone HyG4 and the *Pst* β and *Pst* δ subclones, B. Forget for the plasmids JW151 and JW102, R. A. Flavell, J. Edwards and D. Roberts for helpful discussion, O. Smithies and B. Forget for sending us their results before publication, S. Goodbourn for preparing the *Pst* β and *Pst* δ subclones, and H. Ayyub for technical assistance.

Received 1 December 1980; accepted 9 February 1981.

1. Proudfoot, N. J. & Baralle, F. E. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5435-5439 (1979).
2. Flavell, R. A., Kooter, J. M., DeBoer, E., Little, P. F. R. & Williamson, R. *Cell* **15**, 25-41 (1978).
3. Little, P. F. R., Flavell, R. A., Kooter, J. M., Annison, G. & Williamson, R. *Nature* **278**, 227-231 (1979).
4. Fritsch, E. F., Lawn, R. M. & Maniatis, T. *Nature* **279**, 598-603 (1979).
5. Fritsch, E. F., Lawn, R. M. & Maniatis, T. *Cell* **19**, 959-972 (1980).
6. Jeffreys, A. J., Craig, I. W. & Francke, U. *Nature* **281**, 606-608 (1979).
7. Amin, A. B. *et al. Br. J. Haemat.* **43**, 537-548 (1979).
8. Ottolenghi, S. O. *et al. Nature* **278**, 654-656 (1979).
9. Bernards, R., Kooter, J. M. & Flavell, R. A. *Gene* **6**, 265-280 (1979).
10. Orkin, S. H., Alter, B. P. & Altay, C. J. *clin. Invest.* **64**, 866-869 (1979).
11. Bernards, R. & Flavell, R. A. *Nucleic Acids Res.* **8**, 1521-1534 (1980).
12. Tuan, D., Murnane, M. J., de Riel, J. K. & Forget, B. G. *Nature* **285**, 335-337 (1980).
13. Smithies, O. *Cold Spring Harb. Symp. quant. Biol.* **29**, 309-319 (1964).
14. Chow, L., Kahmann, R. & Kamp, D. J. *molec. Biol.* **113**, 591-609 (1977).
15. Roeder, G. S. & Fink, G. R. *Cell* **21**, 239-240 (1980).
16. Slightom, J. L., Blechl, A. E. & Smithies, O. *Cell* **21**, 627-638 (1980).
17. Lawn, R. M., Efstratiadis, A., O'Connell, C. & Maniatis, T. *Cell* **21**, 647-651 (1980).

18. Spritz, R. A., DeRiel, J. K., Forget, B. G. & Weissman, S. M. *Cell* **21**, 639-646 (1980).
19. Weatherall, D. J. & Clegg, J. B. *The Thalassemia Syndromes* 3rd edn (Blackwell Scientific, Oxford, 1981).
20. Huisman, T. H. J. *et al. Ann. N.Y. Acad. Sci.* **232**, 107-124 (1974).
21. van der Ploeg, L. H. T. *et al. Nature* **283**, 637-642 (1980).
22. Papayannopoulou, T., Kalmantis, T. & Stamatoyannopoulos, G. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6420-6424 (1979).
23. Comi, P. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 362-365 (1980).
24. Jeffreys, A. J. *Cell* **18**, 1-10 (1979).
25. van der Ploeg, L. H. T. & Flavell, R. A. *Cell* **19**, 947-958 (1980).
26. Orkin, S. H., Old, J. M., Weatherall, D. J. & Nathan, D. G. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2400-2404 (1979).
27. Mears, J. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 1222-1226 (1978).
28. Wahl, G. M., Stern, M. & Stark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3683-3687 (1979).
29. Gross-Bellard, M., Oudet, P. & Chambon, P. *Eur. J. Biochem.* **36**, 21-38 (1973).
30. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
31. Kan, Y. W. & Dozy, A. M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5631-5635 (1978).
32. Maniatis, T., Jeffrey, A. & Kleid, D. G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1184-1188 (1975).
33. Orkin, S. H. *et al. Cell* **17**, 33-42 (1979).
34. Laskey, R. A. & Mills, A. D. *FEBS Lett.* **82**, 314-316 (1977).
35. Wilson, J. T. *et al. Nucleic Acids Res.* **5**, 563-581 (1978).

LETTERS

Eternity matters

Don N. Page & M. Randall McKee

Departments of Physics and Astronomy, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

The matter content of an ever-expanding Friedmann universe will never become negligible compared with the radiation, assuming electrons and positrons ($e^\pm$) are absolutely stable, even if all other massive particles are unstable. The possible annihilation of $e^\pm$ in a $k=0$ model occurs so slowly that the radiation density can never greatly exceed the matter density. For a very homogeneous distribution of $e^\pm$ which cool at just the expansion rate, the asymptotic ratio of matter to radiation density in a $k=0$ universe is calculated here to be $(\sqrt{105}-3)/12=0.6039$. Heating by collisional recombination would probably keep the matter density far higher. A certain amount of clumping might make the annihilation occur faster, but larger inhomogeneities are likely to lead to black holes which could be even more effective in keeping the universe from becoming radiation dominated.

Several hypotheses¹⁻⁴ have been proposed to describe the future of an ever-expanding universe. To make some definite predictions, we assume here an open universe that is approximately homogeneous and isotropic (described by a Friedmann model) and which obeys Einstein's equations with no cosmological constant. (Without some such assumptions we could not rule out any conceivable influence coming in from beyond our present particle horizon.) We also assume that charge is absolutely conserved (so $e^\pm$, as the lightest charged particles, are stable) and that no phase transitions bring in fundamentally

new physics at arbitrarily low temperatures. Within such a model, the older picture of stray elementary particles, atoms, stars and black holes spreading out in space has been greatly altered by predictions of the decay of baryons⁵⁻⁸ and black holes⁹. Thus it seems that the only stable particles may be electrons, positrons, neutrinos, photons and gravitons. In a $k=-1$ Friedmann model these particles would spread further and further apart and have a negligible chance of ever meeting after a sufficiently late time, so such a universe would always remain matters dominated. It has been argued³ that a $k=0$ universe would become radiation dominated after the thermal kinetic energy of the electrons and positrons ($e^\pm$) drops below the Coulomb potential energy between pairs so that they become electrically bound and eventually annihilate. However, during this slow annihilation the energy in radiation is being redshifted, so a more careful analysis is needed to compare the decreasing mass density of matter with that of radiation.

Assume a $k=0$ Friedmann universe expands according to $R \propto t^\alpha$ where $1/2 \leq \alpha \leq 2/3$, each extreme representing complete domination by either radiation or matter respectively. The direct collisional annihilation of free $e^\pm$ has a low-velocity cross-section¹⁰ $\sigma \propto v^{-2} \propto t^{2\alpha}$ and hence a probability of occurring which becomes negligible at late times if $\alpha > 1/2$. It is more probable for $e^\pm$ to bind into positronium first. However, two-body radiative 'recombination' has a total cross-section¹¹ only logarithmically greater than the direct collisional annihilation and is, therefore, also negligible. Collisions of neutral particles are even more rare, so any stable massive neutral particles would keep the universe matter dominated. To continue the argument against radiation domination, we set this case aside and assume $e^\pm$ are the only stable massive particles.

Three-body collisional recombination of positronium is the dominant mechanism leading to $e^\pm$ annihilation at late times. Let $N = xt^{-3\alpha}$ be the number density of all $e^\pm$ that have not

annihilated and $N_e = \gamma t^{-3\alpha}$ be that of free e^+ s. Then if the temperature of free e^+ s drops as $T \propto t^{-2\alpha}$,

$$\begin{aligned} t^{-3\alpha} \, dy/dt &= -N_e \sum_n C_{i,n} \\ &\sim -\gamma^3 t^{-9\alpha} (m_e^{-4} e^{-4} T^{-1} n_{\max}^7) \\ &\propto -\gamma^{11/6} t^{-7\alpha/2} \end{aligned} \quad (1)$$

using the collisional ionization cross-section^{12,13} and detailed balance to get the recombination rate coefficient $C_{i,n}$ into the n th level and then summing over all levels up to $n_{\max} \sim m_e^{1/2} e \gamma^{-1/6} t^{\alpha/2}$ where the size of the positronium atom equals the separation between free e^+ s. (There is no problem forming new positronium atoms surrounding smaller ones, because as neutral objects the smaller atoms will have a negligible effect on the larger ones.) Thus at late times the fraction of the original e^+ s that are still free is proportional to $\gamma \propto t^{-3(2-\alpha)/5} \propto n_{\max}^{-6\beta}(t)$, where

$$\beta = (2-\alpha)/(2+4\alpha) \quad (2)$$

After forming in levels around $n_{\max}(t)$, the positronium spirals into the ground state by radiative transitions, annihilating at a time $t_a \propto n_{\max}^6(t) \propto \gamma^{-1/\beta}(t)$. Hence the fraction of e^+ s that have not annihilated by time t_a is proportional to $x(t_a) \approx \gamma(t) \propto t_a^{-\beta}$. This decreases more slowly than the redshifting black-body radiation energy, so the e^+ mass density ρ_e always dominates that radiation density. However, the e^+ annihilation gives a coupling to the total radiation density ρ_r by

$$d\rho_e/dt = (-3H - \beta/t)\rho_e, \quad d\rho_r/dt = -4H\rho_r + (\beta/t)\rho_e \quad (3)$$

In a $k=0$ universe dominated by these two densities so that $H^2 \equiv \dot{R}^2/R^2 = 8\pi(\rho_e + \rho_r)/3$, the asymptotic solution of equation (3) is

$$\begin{aligned} \rho_e &= (1-2\beta)(1-\tfrac{1}{2}\beta)/6\pi t^2, \\ \rho_r &= \tfrac{3}{2}\beta(1-\tfrac{1}{2}\beta)/6\pi t^2, \quad R \propto t^{(2-\beta)/3} \end{aligned} \quad (4)$$

This gives the $R \propto t^\alpha$ expansion assumed above with equations (2) and (4) implying

$$\alpha = (3 + \sqrt{105})/24 = 0.5520, \quad \beta = (13 - \sqrt{105})/8 = 0.3441 \quad (5)$$

Then $\rho_e/\rho_r = (2-4\beta)/3\beta = (\sqrt{105}-3)/12 = 0.6039$. Even though the e^+ s are annihilating, their density is thus never completely dominated by the radiation density.

The annihilation of e^+ s may be even slower than this, because the collisional recombination of positronium will probably keep the temperature of the free e^+ s from dropping as fast as $t^{-2\alpha}$. In fact, if the temperature is kept comparable with the energy that is released when each e^+ pair becomes bound, $T \sim m_e^4 n_{\max}^{-2}$, then equation (1) becomes $t^{-3\alpha} \, dy/dt \propto -\gamma^{3/2} t^{-9\alpha/2}$. This gives $\beta = 1-3\alpha/2$ with the asymptotic solution (4) implying $\alpha = 2/3$, $\beta = 0$. In this case the number of e^+ s decreases only logarithmically with time, and the universe stays completely matter dominated.

All this assumes a nearly perfect Friedmann model with negligible inhomogeneities. Clumping of the e^+ s will make their annihilation occur faster but the annihilation itself will tend to destroy the inhomogeneities. The asymptotic behaviour will depend on the spectrum of the inhomogeneities at large scales and their growth compared with their destruction. Inhomogeneities with vorticity insufficient to prevent them from collapsing gravitationally may form black holes if they are large enough to collapse completely before their e^+ s annihilate. (This requires a mass $M \geq e^4 m_e^{-3} \sim 10^{25} M_\odot$.) If black holes (which are here classified as matter as their rest mass dominates their energy) form with a spectrum¹⁴ $\propto M^{-\gamma} dM$ up to arbitrarily high mass M , their decay⁹ with lifetime $\propto M^3$ will produce a coupling to radiation with $\beta_H = (\gamma-2)/3$. (We still assume an approxi-

mate $k=0$ Friedmann model on sufficiently large scales.) If $\beta_H > \beta$, the black hole density ρ_H will decrease faster than ρ_e and equation (4) will apply. If not, ρ_H will equal or dominate ρ_e and equation (4) will have ρ_e replaced by ρ_H and β replaced by β_H . However, the black holes will probably coalesce into ever bigger and bigger holes. Indeed, even the radiative 'recombination' cross-section $\sigma \sim M^2 v^{-18/7}$ for two black holes to become bound and eventually spiral together will cause each black hole to undergo an infinite number of coalescences (if it gets big enough not to evaporate between collisions) for $\alpha \leq 0.7$, unlike the case for e^+ radiative recombinations which require $\alpha \leq 0.5$ to be effective. Multibody interactions are likely to be even more effective, as gravity is always attractive. Thus black holes may keep a $k=0$ universe always entirely matter dominated.

If, on the other hand, there is a maximum mass to the black holes that form, then they will all evaporate, producing mostly massless radiation. This could make the universe radiation dominated temporarily, but eventually the stray matter or the roughly 10^{44} e^+ s emitted during the final stages of each black hole decay¹⁵ will become comparable with the radiation density. Thus in any case radiation will never completely dominate the density of an ever-expanding approximately Friedmann universe; matter will always be important.

More details of the e^+ annihilation in a $k=0$ universe, including the spectrum of the enormous number of photons produced, will be given elsewhere¹⁶.

D.N.P. appreciates support from NSF grant PHY79-18430.

Received 12 December 1980; accepted 3 March 1981.

1. Davies, P. C. W. *Mon. Not. R. astr. Soc.* **161**, 1-5 (1973).
2. Islam, J. N. Q. *Jl R. astr. Soc.* **18**, 3-8 (1977); *Sky Telescope* **57**, 13-18 (1979).
3. Barrow, J. D. & Tipler, F. J. *Nature* **276**, 453-459 (1978).
4. Dyson, F. J. *Rev. mod. Phys.* **51**, 447-460 (1979).
5. Pati, J. C. & Salam, A. *Phys. Rev.* **D8**, 1240-1251 (1973); *Phys. Rev. Lett.* **31**, 661-664 (1973).
6. Georgi, H. & Glashow, S. L. *Phys. Rev. Lett.* **32**, 438-441 (1974).
7. Georgi, H., Quinn, H. R. & Weinberg, S. *Phys. Rev. Lett.* **33**, 451-454 (1974).
8. Zel'dovich, Ya. B. *Phys. Lett.* **59A**, 254 (1976); *Soviet Phys. JETP* **45**, 9-11 (1977).
9. Hawking, S. W. *Nature* **248**, 30-31 (1974); *Commun. Math. Phys.* **43**, 199-220 (1975).
10. Bransden, B. H. *Case Studies in Atomic Collision I* (eds McDaniel, E. W. & McDowell, M. R. C.) (North-Holland, Amsterdam, 1969).
11. Kramers, H. A. *Phil. Mag.* **46**, 836-871 (1923).
12. Thomson, J. J. *Phil. Mag.* **23**, 449-457 (1912).
13. Percival, I. C. & Richards, D. *Adv. atom. molec. Phys.* **11**, 1-82 (1975).
14. Carr, B. J. *Astrophys. J.* **201**, 1-19 (1975).
15. Page, D. N. *Phys. Rev.* **D16**, 2402-2411 (1977).
16. Page, D. N. & McKee, M. R. Preprint (Pennsylvania State University, 1980).

Ultra-heavy cosmic ray studies with Ariel VI

P. H. Fowler, R. N. F. Walker, M. R. W. Mashedor, R. T. Moses & A. Worley

H. H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK

The satellite Ariel VI was launched on 3 June 1979 into a near-circular 625 km 55° inclination orbit. The Bristol experiment on board is of a novel design using a spherical gas scintillation chamber with an integral plastic Cerenkov element¹. The data presented here were collected during an exposure of $\sim 360 \text{ m}^2 \text{ sr days}$ with essentially 100% recording efficiency for events with $Z \geq 38$. The abundance spectrum, particularly for $Z \geq 50$, suggests that the cosmic ray source is significantly enriched in r -process material relative to the Solar System.

Charge estimates for ultraheavy nuclei ($Z \geq 30$) are made on the assumption that both components of the signal increase as Z^2 and use a daily calibration at ^{26}Fe based on about 15,000 events. Both components of the signal have remained essentially

Table 1 Comparison of the spectrum $N(Z_{\text{est}})$ from the present experiment with those of previous experiments on balloons and Skylab

Charge	Present experiment	Balloon data (top of atmosphere)	Skylab
26	10^6	10^6	10^6
50	4.3 ± 1.4	$*7.5 \pm 2.3$	
52	6.5 ± 1.5	$*6.2 \pm 2.0$	
54	2.2 ± 1.1	$*5.5 \pm 1.7$	
56	7.7 ± 1.5	$*5.4 \pm 1.7$	
58	1.9 ± 1.1	$*5.8 \pm 1.7$	
60	2.2 ± 1.0	$*4.8 \pm 1.4$	
62	1.8 ± 0.8	$*3.1 \pm 1.0$	
64	3.1 ± 0.9	$*1.8 \pm 0.7$	$*3.0 \pm 1.0$
66	2.9 ± 1.0	$*1.2 \pm 0.5$	$*2.7 \pm 0.9$
68	0.1 ± 0.6	0.6 ± 0.4	$*0.8 \pm 0.4$
70	1.5 ± 0.7	0.8 ± 0.3	$*1.6 \pm 0.5$
72	1.3 ± 0.7	1.0 ± 0.3	$*0.9 \pm 0.4$
74	1.8 ± 0.8	1.1 ± 0.3	0.9 ± 0.3
76	0.9 ± 0.6	1.5 ± 0.4	1.6 ± 0.4
78	3.1 ± 0.9	1.8 ± 0.5	1.3 ± 0.4
80	0.9 ± 0.7	1.3 ± 0.4	0.9 ± 0.3
82	1.7 ± 0.8	0.9 ± 0.4	0.6 ± 0.3
84	0.4 ± 0.6	0.8 ± 0.3	0.8 ± 0.3
86	—	0.5 ± 0.3	0.2 ± 0.2
88	0.2 ± 0.2	0.3 ± 0.2	0.0
≥ 90	0.3 ± 0.2	1.1 ± 0.4	0.6 ± 0.3

Based on 137 events $50 < Z < 65$ and 102 events $Z > 65$ 84 events $50 < Z < 65$ and 96 events $Z > 65$ 104 events $Z > 65$

The balloon data have been corrected to yield the fluxes at the top of the atmosphere for a better comparison with the space data. The nominal exposures involved for the present experiment and the balloon data are each $\sim 400 \text{ m}^2 \text{ sr}$ days, and for Skylab $\sim 700 \text{ m}^2 \text{ sr}$ days. The balloon data is all obtained at high geomagnetic latitude and under $\sim 4 \text{ g cm}^{-2}$ of overlying air. Numbers using wider charge bins are also given for high Z because of the correlated errors in our data.

* The value is subject to significant and rapidly charge-dependent corrections due to threshold effects for the tracks in plastic.

constant since launch. The linearity has been confirmed for the lower charges with the observation of peaks for all the even charges $8 \leq Z \leq 26$ and the charge resolution at iron is $\sigma \sim 0.4$ charge. Data collected in the neighbourhood of the South Atlantic Anomaly and during occasional periods of interference from high particle rates in the auroral belts have been rejected.

Figure 1 shows the charge spectrum of apparent charge Z_{app} near ^{26}Fe for events collected in regions where the geomagnetic cutoff is between 3 and 5 GV. The tail above $Z = 26$, caused by the relativistic rise in the scintillation signal, reduces the ^{28}Ni peak to a shoulder and smothers the contribution from ^{30}Zn . This process results in an uncertainty in charge ΔZ which is proportional to Z and is the dominant term at high Z . Nuclear fragmentation within the thin ($\sim 1 \text{ g cm}^{-2}$) detector will produce a tail to low Z_{app} which is masked by the form of the charge spectrum. Additionally, both electron showers generated in the spacecraft body and stopping iron nuclei have the ability to mimic high charge events, but these spurious events can be eliminated by suitable selection of the data. A plastic scintillator guard counter intercepts all electron shower events from the spacecraft and, unavoidably, $\sim 30\%$ of genuine events; all these events must be rejected to remove the showers. Electron shower contamination was found to be sufficiently low for $Z \geq 62$ that all events were accepted regardless of the guard counter signal. Spurious apparently high charge events due to stopping iron group nuclei can be produced by the very small fraction of such nuclei whose entry ranges are both $\geq 7 \text{ g cm}^{-2}$ and exceed by only $\sim 0.1 \text{ g cm}^{-2}$ the total path in the detector, such that an excessively large energy deposition occurs during their passage through the final 0.01 g cm^{-2} of gas. The extent of the excess energy loss does, however, have a well-defined upper limit, and only events with $Z_{\text{app}} \leq 50$ may be produced in this way by ^{26}Fe or ^{28}Ni . For this reason the charge spectrum has been divided at $Z_{\text{app}} = 50$ and high latitude events excluded from the lower charge accumulation.

The spectrum of apparent charge that results from the experiment contains the effects of the relativistic rise. These effects can be removed statistically to give a spectrum of estimated charge Z_{est} . It is assumed that the spectra of energy per

nucleon for ultraheavy cosmic rays and for ^{26}Fe nuclei are indistinguishable². Further evidence for the similarity of the energy spectra is obtained in this experiment from comparison of the flux values as a function of the geomagnetic cutoff. Table 1 compares our spectrum of Z_{est} with results from the Skylab experiment³ and balloon work². The comparison is valid only for $Z \geq 50$ due to the threshold effects of track registration in the Lexan plastic detectors. There are features common to the three spectra, the most significant being the overall flux values, the prominent peak at $Z \sim 78$ and a comparable abundance of nuclei with $58 \leq Z \leq 66$. The Skylab estimate for $Z \sim 66$ was considered by its authors to be uncertain due to the proximity of the threshold referred to above. The general agreement between the sets of results supports the charge scale used in the Lexan track analysis in both the Skylab and balloon experiments.

Table 2 presents the data from this experiment for the whole available charge range. The spectra of Z_{app} and Z_{est} are given together with the original number of events on which each entry was based. Also shown is our estimate of the charge spectrum of secondaries produced by interactions in the interstellar medium that are included among the nuclei detected. The difference between the two figures is the estimated number of residual primaries. This is extrapolated back to the source region to give our estimated source spectrum which again is normalized to $\text{Fe} = 10^6$ and which can be directly compared with the Solar System abundances compiled by Cameron⁴. We used the simplest form of 'leaky box' model for cosmic ray propagation which ignores ionization energy loss and assumes all cross-sections for interaction with interstellar hydrogen to be independent of energy. Cross-sections for nuclear interactions were based on those given for 2.3 GeV per nucleon by R. Silberberg (personal communication). The escape mean free path was taken as 5.0 g cm^{-2} of hydrogen. Clearly, except for the highest charges with $Z \geq 78$, a substantial fraction of the observed ultraheavy nuclei are secondaries arising from interactions in interstellar space. This will be the case whenever the amount of matter traversed by the detected particles is comparable with their interaction mean free paths.

Several authors have concluded that the ultraheavy cosmic ray spectrum seems to be enriched in material which has undergone nucleosynthesis by the rapid neutron capture process. More recently it has been suggested⁵ that at least for the elements $Z \leq 28$ the cosmic ray source spectrum could be simply explained as a sample of normal interstellar material, modified only by the effects of differing ionization potentials for the various elements. The cosmic ray source spectrum in the ultra-heavy domain would on this assumption have much in common

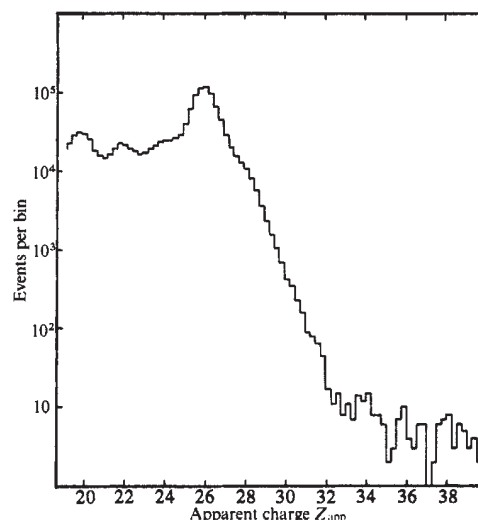
**Fig. 1** Histogram of apparent charge of events collected during 110 days of operation for cutoffs between 3GV and 5GV. The sample contains $\sim 7 \times 10^5$ iron nuclei.

Table 2 Estimates of the spectra at several stages between detection and the computed source composition together with the Solar System composition of Cameron

Charge	Present experiment				Solar System abundances
	$N(Z_{app})$	$N(Z_{est})$	N_{sec}	N_{source}	
Fe = 26	6.3×10^6	6.3×10^6	6.3×10^6	10^6	10^6
34	557 (213)	526	217	59 ± 9	83
36	327 (125)	290	145	29 ± 7	54
38	298 (114)	301	97	41 ± 7	31
40	196 (74)	175	58	24 ± 6	17
42	93 (35)	71	43	6 ± 4	4.9
44	66 (25)	63	37	5.7 ± 3.8	2.3
46	42 (16)	36	27	2.1 ± 3.1	1.9
48	48 (18)	51	26	6.0 ± 3.5	2.1
50	32 (26)	27	32	-1.2 ± 2.1	4.4
52	38 (31)	41	19.6	5.3 ± 2.3	8.1
54	20 (16)	13.8	20	-1.7 ± 1.8	7.4
56	40 (33)	48	20	7.5 ± 2.5	5.8
58	20 (16)	12.0	6.6	1.5 ± 1.9	1.6
60	15 (12)	14.0	10.0	1.1 ± 1.7	0.98
62	12	11.1	13.3	-0.6 ± 1.4	0.32
64	17	19.4	12.5	2.0 ± 1.7	0.56
66	18	18.5	6.8	3.5 ± 1.8	0.50
68	6	0.8	6.0	-1.6 ± 1.1	0.33
70	8	9.3	5.9	1.1 ± 1.3	0.26
72	8	8.1	6.7	0.4 ± 1.3	0.22
74	10	11.1	3.9	2.3 ± 1.5	0.37
76	7	5.6	5.4	0.1 ± 1.3	1.20
78	15	19.3	3.4	5.2 ± 2.0	2.08
80	9	6.0	2.5	1.2 ± 1.6	0.46
82	10	10.7	0.9	3.4 ± 1.7	3.07
84	5	2.4	0.3	0.7 ± 1.2	0.08
86	3	—	—	—	—
88	1	1	—	0.3 ± 0.3	—
≥ 90	2	2	—	0.7 ± 0.5	0.08

For the highest charges all the detected events were used. For $Z \leq 60$, where selections were necessary, the values in brackets refer to the number of events utilized. The errors shown are due to Poisson statistics alone and are quite strongly negatively correlated in adjacent charge bins. Wider bins are therefore used at higher charges to display the features of the charge spectrum.

with the Solar System composition at all charges and, indeed, the overall abundance in our source spectrum is very similar to the Solar System value, a point which has been apparent from previous experiments for several years. In addition, the main features of the Solar System composition, namely the peaks at $Z \sim 54$ and $Z \sim 80$, are both clearly present in our data. However, the abundance ratio between $52 \leq Z \leq 56$ and $74 \leq Z \leq 84$ is 0.86 ± 0.30 compared with a Solar System value of 2.9, and the predominant species in the upper peak is ^{78}Pt rather than ^{82}Pb . A striking anomaly also occurs at $Z \sim 64$ where the cosmic rays seem to be ~ 4 times over-abundant. These discrepancies cannot be attributed to the effects of ionization potentials. If as an alternative we suppose that much shorter mean free paths were to apply in the leaky box, the estimated source spectrum would approach the detected spectrum thereby reducing the excess for $Z \geq 74$ in the cosmic ray source. At the same time, however, the rare earths as a whole ($58 \leq Z \leq 72$) would become severely overabundant and the overall similarity to the Solar System abundances would not be improved.

There is a small Solar System abundance anomaly at $Z \sim 64$ ($A \sim 164$) which has been attributed to fission fragments. Steinberg and Wilkins⁶ consider that the feature may be produced by the heavier fragment resulting from asymmetric fission of superheavy nuclei with $A \sim 300$. Such nuclei are produced in nucleosynthesis models involving a cyclic r -process, but will be accompanied by substantially greater quantities of $A \sim 280$ nuclei. Our results seem inconsistent with this interpretation as our source spectrum contains neither a feature at $A \sim 280$ nor the excessive abundance at $A \sim 134$ ($Z \sim 54$) which would result from the symmetric fission of the $A \sim 280$ nuclei advocated by Steinberg and Wilkins. We propose as an alternative that the $A \sim 280$ feature itself is responsible, by asymmetric fission, for the anomaly at $Z \sim 64$ in both the cosmic ray and Solar System abundances, the original mass being distributed between two peaks at $A = 105$ and $A = 164$ and per-

haps 10 neutrons. Our data have as yet insufficient statistical weight to determine whether the region of the lighter peak, at $Z \sim 44$, is appropriately enhanced.

In the charge region $Z \geq 90$ we would have expected, on the basis of the Skylab and Solar System abundances, to have recorded respectively ~ 4 and ~ 0.2 such events in the present sample. Of the two events observed, the event with $Z = 99$ could be a perfectly normal actinide. Its light was well distributed among the photomultipliers of the detector sphere and the absence of a plastic scintillator signal rules out a shower origin in the spacecraft body. Because the highest charge event, with $Z = 114$, did activate the guard counter its interpretation as an electron shower cannot be excluded. Our conclusion that the cosmic ray source seems to be significantly enriched in the heavier r -process nuclei therefore rests on our discussion of the charge region $52 \leq Z \leq 84$.

Received 19 February; accepted 23 March 1981.

1. Fowler, P. H. *et al. Proc. 16th Int. Cosmic Ray Conf.* **12**, 338 (1979).
2. Fowler, P. H. *et al. Proc. 15th Int. Cosmic Ray Conf.* **11**, 165 (1977).
3. Shirk, E. K. & Price, P. B. *Astrophys. J.* **220**, 719 (1978).
4. Cameron, A. G. W. Preprint no. 1357 (Harvard-Smithsonian Center for Astrophysics, 1980).
5. Cassé, P. & Goret, P. *Astrophys. J.* **221**, 703 (1978).
6. Steinberg, E. P. & Wilkins, B. D. *Astrophys. J.* **223**, 1000 (1978).

Albedo change by man: test of climatic effects

Gerald L. Potter, Hugh W. Ellsaesser,
Michael C. MacCracken & James S. Ellis

Lawrence Livermore National Laboratory, University of California,
PO Box 808 Livermore, California 94550, USA

Sagan *et al.*¹ have suggested that the cumulative impact of anthropogenic albedo changes may have contributed to global climate changes in the past and that its effect may be continuing. Using a statistical dynamic climate model (SDM) with more realistic surface albedo changes than previously used², we have computed the combined impact of desertification of the Sahara and deforestation of the tropical rain forest. While the model computed a surface cooling of 0.6 K for the Northern Hemisphere, the global mean of ~ 0.2 K was substantially less than the 1 K suggested by Sagan *et al.*¹. We infer that man's cumulative impact on planetary surface albedo over the past few thousand years has had a small and probably undetectable effect on global climate.

Earlier numerical model experiments²⁻⁴, suggesting that the main effect of man's past activities should have been a general increase of surface albedo resulting in lower surface temperatures, were not designed to incorporate more than single impacts, such as tropical deforestation or desertification. The suggestion of Sagan *et al.*¹ that combined effects may have long term global ramifications seemed sufficiently serious to warrant investigation in a more sophisticated model. We have therefore used the Lawrence Livermore National Laboratory (LLNL) two-dimensional SDM to calculate the combined climatic effect of desertification in the Northern Hemisphere subtropics and deforestation of the tropical rain forest. Other surface modifications suggested by Sagan *et al.*¹ were not included as they involved relatively small surface areas and there is less agreement as to their potential impact on climate.

The LLNL two-dimensional SDM model⁵ is similar in many ways to general circulation models using the basic conservation equations to calculate zonal mean fields of pressure, temperature, wind and humidity. The model has nine pressure layers and a 10° -latitude grid. The surface area of each latitude zone is divided into land (of differing types and elevations) and ocean (partially covered by ice). Sea ice (with the appropriate albedo) is formed when the ocean temperature drops below 271.4 K. New ice spreads at a minimum thickness of 20 cm until it

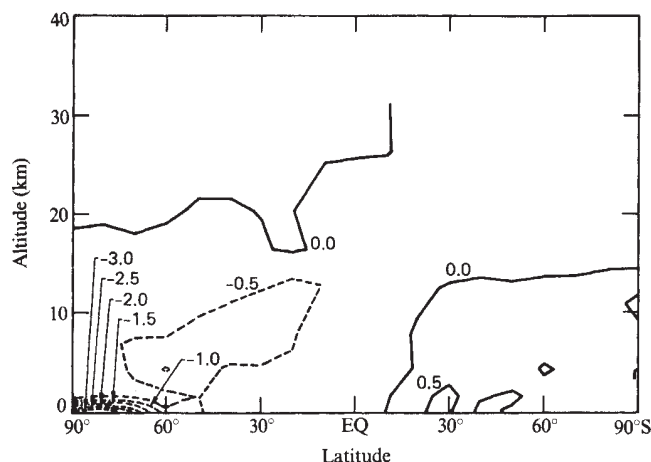


Fig. 1 Latitude-height temperature difference (perturbed minus control). Contour interval = 0.5 K.

occupies the maximum allowable fractional coverage (leaving some open ocean to simulate open leads). If the energy deficit continues the sea ice then increases in depth. Conversely, an energy excess would melt the ice before the water warms. The melt water (with the appropriate albedo) can refreeze on top of the older ice and thus increase the ice depth. Fractional grid zone assignments of various surface types with corresponding physical characteristics (for example, albedo, heat capacity, bulk transfer coefficients) enable the sensitivity of the model to modifications of these parameters to be tested.

Surface fluxes of sensible and latent heat are calculated using the vertical gradients of water vapour and temperature, the surface wind velocity and low-level convective activity. Clouds are randomly correlated vertically and are a prescribed latitude and height dependent function of relative humidity⁶ (for further details see refs 4 and 7).

The sensitivity, or response to an imposed perturbation, of the model is conveniently measured by its response to an increase of its solar constant. The usual unit of comparison is the sensitivity parameter β (equal to 100 times the temperature change resulting from a 1% increase in solar flux), and experiments with our SDM give a global sensitivity factor $\beta = 70$ K. This is less than half the value quoted for simple energy balance models^{8,9} and the Geophysical Fluid Dynamics Laboratory sector model¹⁰. At least some of this reduced sensitivity results from use of a global model and from the greatly reduced sensitivity of the Southern Hemisphere, which exhibits significantly less ice-albedo and land-surface albedo feedback.

The non-seasonal mode of the LLNL SDM simulates the annual average climate by using a fixed solar zenith angle for each latitude calculated to supply the appropriate annual average solar flux. By eliminating the diurnal and seasonal fluctuations, the model approaches a pseudo-equilibrium state after ~200 days of model integration time. This reflects the adjustment time implied by the heat capacities of the atmosphere, land, snow, ice and arbitrarily thin oceanic mixed layer. For this experiment, two model equilibrium states were achieved: a control case with standard surface conditions corresponding to a pristine environment; a perturbed case corresponding to the suggestion of Sagan *et al.*¹ in terms of tropical deforestation and desert denudation (see Table 1). The surface bulk transfer coefficients, the thermal and hydrologic characteristics were not altered; the only surface parameter modified for those areas perturbed was surface albedo which, after modification, was held constant. This reduced the degrees of freedom for the SDM.

The response to this change in albedo was generally consistent with previous surface albedo modification sensitivity studies^{2,3}. Figure 1 shows the temperature difference (perturbed minus control) between the two cases. Generally, the Northern

Table 1 Model albedo changes from Sagan *et al.*¹

Process	Land type change	Area [km ²] (% band)	Centred on latitude bands	Surface albedo change
Desertification	savanna to desert	9×10^6 (22%)	20° N	0.16 → 0.35
Tropical deforestation	forest to field savanna	7×10^6 (5% of 10° S, 10% of Equator)	10° S and Equator	0.07 → 0.16

Hemisphere troposphere cooled, with a maximum change due to ice-albedo feedback near the surface at 70° N.

There seem to be two important factors in the chain of events leading from surface modification to cooling in the Northern Hemisphere. The first is the direct reduction in surface absorption in the zones of increased surface albedo (see Table 1). The change at 20° N was sufficiently strong to strengthen the temperature gradient between the Sahara³ and the Equator which caused the northern Hadley cell to strengthen and shift southwards. This in turn weakened the hydrological activity at 20° N with less precipitable water associated convection and latent heat release, accompanied by maximum cooling in the upper troposphere, lowering of the tropopause and warming of the lower stratosphere. With less solar energy absorbed in the subtropics there was reduced poleward transport of sensible and latent heat and a general cooling of the higher northern latitudes. The second factor cooling the Northern Hemisphere was the ice-albedo feedback due to increased sea-ice coverage, primarily at 70° N.

Generally, the cloud cover increased in the Northern Hemisphere where the surface temperature had cooled 0.6 K. This agrees with reports¹¹⁻¹³ of global increases in cloud amount with decreased surface temperatures (Table 2). One exception was a general cloud decrease at 20° N due to increased Hadley subsidence. This latitude received the largest surface-albedo perturbation and hence the highest directly resultant surface temperature decrease¹⁴.

In the Southern Hemisphere, only a small fraction of the latitude band at 10° S also had an albedo increase from deforestation. Figure 2 shows the zonal average conversion from potential to kinetic energy due to vertical motion. The heated air in the low level equatorial Hadley cell rises converting potential to kinetic energy. In the higher latitudes the sinking motion converts back to potential energy. This conversion term gives an

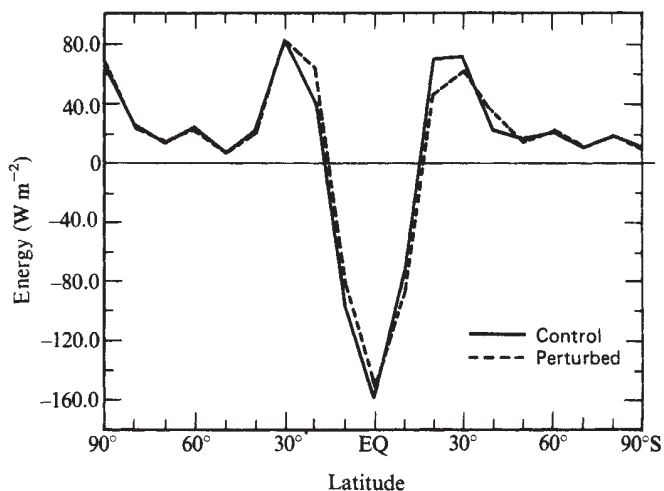


Fig. 2 Kinetic to potential energy conversion through vertical motion.

indication of the intensity of the Hadley cell. In the perturbed case, the Hadley cell shifted southwards with the ascending branch weakening slightly and the southern descending branch weakening more. The southward shift brought increased ascent to the southern subtropics. This with the weakened descending branch permitted increased convective activity, leading to clouds with more vertical structure and less horizontal extent. The total cloud coverage thus decreased at 30°S, allowing increased insolation. This increase, along with the albedo increase at the Equator and 10°S, weakened the subtropical temperature gradient and presumably further weakened the Southern Hemisphere Hadley cell. However, other than the additional solar flux at 30°S, there seems little reason for the slight warming of the Southern Hemisphere.

To assess more fully the relative importance of the various model responses to the imposed increase in surface albedo, we can examine the contribution to the change in solar (S) and outgoing longwave (F) radiation at the top of the atmosphere as given by

$$\delta S = \delta S_r + \delta S_{A_c} + \delta S_{\alpha_s} + \delta S_f$$

$$\delta F = \delta F_T + \delta F_r + \delta F_{A_c} + \delta F_f$$

where r , T , A_c , α_s and f identify components due to precipitable water, temperature, cloud amount, surface albedo and fractional change in ice coverage, respectively. These individual contributions have been calculated by recomputing the planetary radiation balance keeping all of the input data fields fixed (control case) except for the perturbed data field of interest (for example, the temperature, water vapour, clouds or surface albedo from the modified albedo case).

Table 3 shows the change in solar, long-wave and net radiation due to the separated changes in various contributing factors; the two major components in changing the planetary energy balance are clouds and surface albedo. The mean global cloud fraction decreased 0.011, resulting in an increase in long-wave loss to space and an increase in solar radiation absorbed by the Earth-atmosphere system. The surface albedo increased as a result of both the imposed albedo modification and the ice-albedo feedback, decreasing the solar radiation absorbed by the previously clear portion of the Earth-atmosphere system. The mean surface temperature decrease reduced the long-wave loss to space because of the cooler radiating temperature.

Surprisingly, water vapour was the least important component in the long-wave response, although it did produce a reduction in long-wave loss to space. This small influence may be because water vapour experienced both vertical and horizontal redistributions in conjunction with the hemispheric asymmetries of the total model response. Thus the change in global mean precipitable water vapour was so small (−1.6%) that the radiative response was nearly insignificant.

As the model calculates sea ice, the ice cover fraction will differ from experiment to experiment. The effect of this change alone (all other effects held fixed) on the global energy balance is shown in Table 3a. Sea ice increased ~14% in the latitude band 65–75°N, which corresponds to an added sea-ice area of $2 \times 10^6 \text{ km}^2$. The 0.26 W m^{-2} decrease in solar radiation represents

Table 3 a Global mean change in solar, outgoing long-wave, and net radiation (δS , δF , and $\delta S - \delta F$) in W m^{-2} of the control model due to insertion of only the perturbed field

	T	r	A_c	α_s	f
δS	—	0.01	1.25	−3.87	−0.26
δF	−0.53	−0.07	0.31	—	−0.03
$(\delta S - \delta F)$	0.53	0.08	0.94	−3.87	−0.23

b Global change in various parameters (perturbed minus control)

Parameter (ψ)	Change	$\Delta\psi/\psi$ (%)
Global surface temperature (T)	−0.2 K	
Mean precipitable water vapour (r)	−0.04 g cm^{-2}	(−1.6%)
Global cloud cover fraction (A_c)	−0.011	(−1.9%)
Surface albedo (α_s)	+0.018	(+8.8%)
Sea ice (f)	$+2 \times 10^6 \text{ km}^2$	(+13.5% in 67–75°N band)

the effect of sea ice alone, without cloud cover or surface albedo contributions. Long-wave radiation responded because of the larger area radiating at a colder sea-ice temperature (reducing the long-wave flux to space).

The sensitivity of the LLNL SDM to the combined effect of deforestation in the tropics and desertification in the subtropics reveals several interesting responses. The Hadley cell shifted southwards and the northern descending branch intensified. Conversely, the southern descending branch decreased in intensity, thereby reducing subsidence. This reduction in stability in the Southern Hemisphere produced clouds with more vertical and less horizontal extent, allowing more solar radiation to reach the Earth's surface and to increase temperatures at 30°S. The Northern Hemisphere temperature decreased because of the reduced solar radiation reaching the surface coupled with additional cooling due to ice-albedo feedback.

While the LLNL SDM computed a Northern Hemisphere surface cooling of 0.6 K, the global cooling of 0.2 K was substantially less than the 1 K suggested by Sagan *et al.*¹ for the cumulative effect of anthropogenic surface modifications over the past several thousand years. Our model results suggest, therefore, that the global temperature response is significantly less and the Northern Hemisphere temperature response slightly less than that of the simple energy balance model cited by Sagan *et al.*¹

We thank Dr W. J. Quirk for suggestions and comments, Connie S. Mitchell for helping with the calculations, Floy Weichel and Joni Schuld for typing the manuscript. This work was performed under the auspices of the US Department of Energy by the Lawrence Livermore National Laboratory under contract no. W-7405-Eng-48. This paper is a continuation of work presented at the International Workshop on Energy-Climate Interactions, Munster, West Germany in March 1980.

Received 15 September 1980; accepted 25 February 1981.

Table 2 Relationship between change in cloud cover and change in surface temperature in this experiment and those of Schneider *et al.*¹¹ (Northern Hemisphere only)

	δT	δA_c	$\delta A_c/\delta T$
LLNL SDM NH 4km cloud	−0.58	+0.004	−0.007
NCAR GCM 3 km cloud	−2.52	+0.015	−0.006

- Sagan, C., Toon, O. B. & Pollack, J. B. *Science* **206**, 1363 (1979).
- Charney, J. G., Stone, P. H. & Quirk, W. J. *Science* **187**, 434 (1975).
- Ellsaesser, H. W., MacCracken, M. C., Potter, G. L. & Luther, F. M. *Q. J. R. met. Soc.* **102**, 655 (1976).
- Potter, G. L., Ellsaesser, H. W., MacCracken, M. C. & Luther, F. M. *Nature* **258**, 697 (1975).
- MacCracken, M. C. thesis, Univ. California (1968).
- London, J. A. *Study of the Atmospheric Heat Balance. Final Rep.* Contract AF19(122)-165 (Research Division, College of Engineering, New York University, 1957).
- Potter, G. L., Ellsaesser, H. W., MacCracken, M. C. & Luther, F. M. in *GARP Ser. Publ. No.* 22 (WMO, Geneva, 1979).
- Cess, R. D. *J. atmos. Sci.* **33**, 1831 (1976).
- Budyko, M. I. *Climate and Life* (Academic, New York, 1974).
- Ramanathan, V. *J. atmos. Sci.* **34**, 1855 (1977).
- Schneider, S. H., Washington, W. M. & Chervin, R. M. *J. atmos. Sci.* **35**, 2207 (1978).
- Henderson-Sellers, A. *Nature* **279**, 786 (1979).
- Roads, J. O. J. *J. atmos. Sci.* **35**, 753 (1978).
- Bjerknes, J., Allison, L. J., Kreins, E. R., Godshall, F. A. & Warnecke, G. *Bull. Am. met. Soc.* **50**, 313 (1969).

Oxidative power of Mn(IV) and Fe(III) oxides with respect to As(III) in terrestrial and aquatic environments

D. W. Oscarson*, P. M. Huang*, C. Defosse† & A. Herbillon†

* Saskatchewan Institute of Pedology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 0W0

† Université Catholique de Louvain, Groupe de Physico-Chimie, Minérale et de Catalyse, Louvain-la-Neuve 1348, Belgium

The biological availability and the toxicological effects of As depend on its chemical state^{1,2}. It is therefore important to know the speciation and transformations of As in terrestrial and aquatic environments. Arsenite, As(III), is much more toxic than arsenate As(v) (ref. 2). The suspended and bottom sediments can potentially detoxify As(III) that enters aquatic systems by converting it to As(v) through abiotic oxidation³. We hypothesized that the oxides of Mn(IV) and Fe(III) may be the primary electron acceptors in the oxidation of As(III) because Mn and Fe readily participate in many oxidation-reduction reactions in natural environments; however, their role in the oxidation of As(III) to As(v) is yet to be established. We report here that Mn(IV) oxide is a very effective oxidant with respect to As(III); on the other hand, despite the thermodynamic favourability³, the evidence obtained from colorimetry and X-ray photoelectron spectroscopy shows that a redox reaction between Fe(III) oxide and As(III) does not occur within 72 h, indicating that the kinetics of the redox reaction between As(III) and Fe(III) is relatively slow.

Birnessite (δ -MnO₂) was prepared and used in this study because it is one of the most common Mn oxides in sediments and soils⁴. The appearance of As(v) in solution after adding solutions of various concentrations of As(III) to Mn(IV) oxide (Table 1) clearly shows that As(III) is converted to As(v) by Mn(IV); in a control experiment, no detectable As(III) was oxidized in the absence of Mn(IV) oxide. With increasing additions of As(III) above 300 $\mu\text{g ml}^{-1}$, the amount of As(v) produced increased very little (Table 1), indicating that the oxidation capacity of the 100 mg sample of Mn(IV) oxide for As(III) was being approached. Assuming that the Mn(IV) oxide (birnessite) contains 9.5% K and the oxygen : manganese ratio is close to 2 (ref. 4), the 216 $\mu\text{g As(v) ml}^{-1}$ in solution after adding 1,000 $\mu\text{g As(III) ml}^{-1}$ to 100 mg of Mn(IV) oxide (Table 1) indicates that at least ~19% of the Mn(IV) initially present is reduced to Mn(II).

Mn(II) is more soluble than Mn(IV) (ref. 5), therefore the high concentrations of Mn and K in solution in the As(III)-Mn(IV)

oxide systems relative to the As(v)-Mn(IV) oxide system (Table 1) is further evidence that Mn(IV) is reduced to Mn(II) by As(III). The reduction of Mn(IV) and the subsequent partial dissolution of the Mn oxide on the addition of As(III) disrupts the crystal structure at the surface of the oxide and thus a portion of the K present within the structure is free to pass into solution. As the Mn oxide contains ~9.5% K (ref. 4), the 90.9 $\mu\text{g K ml}^{-1}$ in solution for the 1,000 $\mu\text{g As(III) ml}^{-1}$ treatment (Table 1) represents ~67% of the K initially present in the Mn oxide sample. The 50 $\mu\text{g K ml}^{-1}$ in solution in the As(v)-Mn oxide system (Table 1) is attributed to exchangeable K present at or near the surface of the oxide. The K in solution in the As(III)-Mn oxide systems is thus attributed to exchangeable plus structural K.

The decrease in the Mn concentration in solution with the increase in the initial As(III) concentration from 300 to 1,000 $\mu\text{g ml}^{-1}$ (Table 1) is attributed to an increase in the formation of a sparingly soluble manganese-arsenate complex, $\text{Mn}_3(\text{AsO}_4)_2$. This interpretation is substantiated by the fact that the solubility product of $\text{Mn}_3(\text{AsO}_4)_2$ (ref. 6) is very much lower than the ion product, $(\text{Mn}^{2+})^3(\text{AsO}_4^{3-})^2$, of the equilibrated solutions in the As(III)-Mn(IV) oxide systems.

Manganese oxide sorbs little As(v) (Table 1). The point of zero charge of birnessite is close to pH 2 (ref. 7), and therefore the oxide has a high negative charge density at pH values used in this study; As(v) (largely H_2AsO_4^- and HAsO_4^{2-}) is also negatively charged at the pH of this system⁸, consequently a high energy barrier must be overcome for As(v) to sorb onto Mn oxide.

When freshwater lake sediments from southern Saskatchewan, Canada were treated with hydroxylamine hydrochloride or sodium acetate (effective extractants for the selective removal of Mn from sediments and soils^{9,10}), the oxidation of As(III) to As(v) by the treated sediments was greatly decreased relative to untreated samples of the sediments¹¹. This is supporting evidence that Mn oxide is a primary sediment component responsible for the oxidation of As(III). In sediments and soils Mn oxides occur as discrete particles and/or as coatings on other sediment/soil components¹²; and Mn is present in both the colloidal and non-colloidal particle size fractions of stream and lake sediments¹⁰.

No detectable As(v) was found in solution after the addition of As(III) to Fe(III) oxide (Table 1), showing that Fe(III) oxide does not oxidize As(III) at neutral pH values within 72 h (earlier studies³ showed that 48 h was sufficient time for 0.7 mg As(III) to disappear when added to 1 g of sediment). The data in Table 1 also show that As(III) is more strongly sorbed by Fe(III) oxide than is As(v); this is attributed to structural and charge differences between the two As complexes: As(III) (H_3AsO_3 , coordination geometry is pyramidal, at pH values <9.2 arsenious acid is predominantly a neutral complex⁸) and As(v) (H_3AsO_4 , coordination geometry is tetrahedral, at pH values >2.3 arsenic acid is largely dissociated and negatively charged⁸).

Table 1 Oxidation of As(III) and sorption of As by Mn(IV) and Fe(III) oxides

Mn(IV) oxide						Fe(III) oxide			
As(III) or As(V) added ($\mu\text{g ml}^{-1}$)	As(III)	As(V) ($\mu\text{g ml}^{-1}$ in solution)	Mn	K	Final pH	As(III) or As(V) added ($\mu\text{g ml}^{-1}$)	As(III) ($\mu\text{g ml}^{-1}$ in solution)	As(V)	Final pH
100 As(III)	ND	83.5 $\pm$ 1.4*	0.41 $\pm$ 0.12	60.5 $\pm$ 0.0	7.1	100 As(III)	3.67 $\pm$ 0.24	ND	6.9
300 As(III)	63.2 $\pm$ 7.0	186 $\pm$ 5	8.08 $\pm$ 0.36	90.0 $\pm$ 1.5	7.1	100 As(IV)	ND	53.6 $\pm$ 1.1	7.7
500 As(III)	213 $\pm$ 4	205 $\pm$ 1	6.06 $\pm$ 0.60	90.5 $\pm$ 0.6	7.3				
1000 As(III)	665 $\pm$ 5	216 $\pm$ 4	4.61 $\pm$ 0.86	90.9 $\pm$ 0.8	7.5				
300 As(V)	ND	298 $\pm$ 1	0.06 $\pm$ 0.02	50.0 $\pm$ 0.9	7.5				

The Mn(IV) (birnessite) and Fe(III) oxides were prepared as outlined elsewhere^{4,15}. 100 mg of Mn(IV) or Fe(III) oxide was suspended in 70 ml of a solution containing the various concentrations of As(III) and As(v) as NaAsO_2 and $\text{Na}_2\text{HAsO}_4 \cdot 7\text{H}_2\text{O}$, respectively. The pH of the suspensions was adjusted to 7.5, and the flasks containing the suspensions were placed in an oscillating shaker at $25 \pm 0.2^\circ\text{C}$ for 72 h. After the reaction period, the concentration of As(III) and As(v) in solution was determined colorimetrically as described elsewhere³.

* Mean $\pm$ s.d.; n = 3. ND, not detectable.

Table 2 Binding energies for As and Fe(III) reference oxides and As(ads)-Fe(III) oxide

Sample	As _{3p_{3/2}}	As _{3p_{1/2}}	Δ^*	Fe _{2p_{3/2}}
NaAsO ₂	143.4	148.2	4.8	—
Na ₂ HAsO ₄ ·7H ₂ O	144.1	148.9	4.8	—
Fe(III) oxide				711.7
As(III)†-Fe(III) oxide	143.7	148.4	4.7	711.3
As(V)†-Fe(III) oxide	144.4	149.2	4.8	711.5

The measurements were performed on a Vacuum Generators ESCA 3 Spectrometer equipped with a Tracor Northern NS 570 signal averager for signal-to-noise ratio enhancement. The samples were ground and dusted on double-sided adhesive tape. The sequence of spectra was such as to embrace the As_{3p} lines by the C_{1s} reference line at 285.0 eV. Each sample was run twice and the precision of the binding energies was ± 0.1 eV. The signal-to-noise ratio was 50:1 for the C_{1s} and Fe_{2p} lines and at least 15:1 for the As_{3p} doublet.

* BE (As_{3p_{1/2}} - As_{3p_{3/2}}).

† Initial concentration of As(III) and As(V) was 100 $\mu\text{g ml}^{-1}$ and the concentration of As(III) and As(V) remaining in solution after the 72-h reaction period is given in Table 1.

About 96% of the As(III) added to the Fe(III) oxide suspension is sorbed onto the surface of the oxide, and the colorimetric method yields no information on the oxidation state of the surface-sorbed As. X-ray photoelectron spectroscopy (XPS) is a powerful tool for investigating the oxidation state of surface-sorbed species. The results of an XPS study of the As-Fe(III) oxide system is presented in Table 2. The absolute use of binding energies alone for the assignment of the oxidation state is generally not reliable, especially in the case of an adsorbed species when the reference used is a bulk compound (in this case NaAsO₂ and Na₂HAsO₄·7H₂O). However, the constancy of the 0.7-eV shift observed between the As(III) and As(V) samples (As_{3p_{3/2}} and As_{3p_{1/2}} lines) on one hand and between the As(III)-Fe(III) oxide and As(V)-Fe(III) oxide samples on the other (Table 2) is strong evidence that As(III) and As(V) are sorbed as such and no change in oxidation state occurs. The 0.3-eV increase when either As(III) or As(V) is adsorbed onto iron oxide reflects extra atomic relaxation and/or Madelung potential effects. The inability of Fe(III) oxide to convert As(III) to As(V) within 72 h despite the thermodynamic favourability indicates that the kinetics of the redox reaction is relatively slow at pH values near neutrality.

The biological methylation of As, Hg and Se has attracted a great deal of attention in recent years. Because the methylation of inorganic arsenicals is essentially a reduction reaction¹³, the abiotic oxidation of As(III) to As(V) by freshwater lake sediments would tend to counteract the methylation process. This is significant because some organic arsenicals such as the various methyl-substituted arsines are more toxic than the inorganic arsenic species in higher oxidation states¹⁴.

Our data indicate that Fe(III) oxide does not convert As(III) to As(V) within 72 h. On the other hand, Mn(IV) oxide, which is common in freshwater sediments, is a very effective oxidant with respect to As(III). Our data thus indicate that Mn(IV) oxide may be very important in natural systems in decreasing the concentration of As(III), a highly toxic pollutant.

This study was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

Received 6 February; accepted 6 March 1981.

- Webb, J. L. *Enzyme and Metabolic Inhibitors* Vol. 3, Ch. 6 (Academic, New York, 1966).
- Ferguson, J. F. & Gavis, J. W. *Res.* **6**, 1259-1274 (1972).
- Oscarson, D. W., Huang, P. M. & Liaw, W. K. *J. Envir. Qual.* **9**, 700-703 (1980).
- McKenzie, R. M. *Miner. Mag.* **38**, 493-502 (1971).
- Stumm, W. & Morgan, J. J. *Aquatic Chemistry*, 547-553 (Wiley-Interscience, New York, 1970).
- Hess, R. E. & Blanchard, R. W. *J. Soil Sci. Soc. Am.* **40**, 847-852 (1976).
- Healy, T. W., Herring, A. P. & Fuerstenau, D. W. *J. Colloid Interface Sci.* **21**, 435-444 (1966).
- Weast, R. C. (ed.) *CRC Handbook of Chemistry and Physics* 58th edn, D-151 (CRC, West Palm Beach, 1978).
- Chao, T. T. *Proc. Soil Sci. Soc. Am.* **26**, 764-768 (1972).

- Oscarson, D. W., Rogers, J. S., Huang, P. M. & Liaw, W. K. *Int. Rev. ges. Hydrobiol.* **66**, 95-107 (1981).
- Oscarson, D. W., Huang, P. M. & Liaw, W. K. *Clays Clay Miner.* (in the press).
- Jenne, E. A. *Am. Chem. Soc., Adv. Chem. Ser.* **73**, 337-387 (1968).
- Braman, R. S. in *Arsenical Pesticides* (ed. Woolson, E. A.) 108-123 (American Chemical Society, Washington DC, 1975).
- Penrose, W. R. *CRC Crit. Rev. Envir. Control.* **4**, 465-482 (1974).
- Huang, P. M., Wang, T. S. C., Wang, M. K., Wu, M. H. & Hsu, N. W. *Soil Sci.* **123**, 213-219 (1977).

Earthquake-induced ground acceleration

Michele Caputo

Istituto di Fisica, G. Marconi, University of Rome,
Piazzale Aldo Moro, 2, I-00185 Rome, Italy

Earthquakes may be characterized after the event by several physical parameters, including the linear dimension of the fault, l , and the stress drop, p . Although the direct measurement of these parameters is often difficult, there are data on the frequency distribution of l and p for certain series of earthquakes (refs 1-11; C. Archambeau, unpublished data; personal communications from T. Mikumo, T. Utsu and R. I. Archuleta). For p there is evidence¹²⁻¹⁴ that the frequency distribution $N(p)$ is given by a relationship of the form

$$\log N(p) = c_1 - (1 - \alpha) \log p \quad (1)$$

where c_1 and α are constants (α lying between 0 and 1) and $N(p) dp$ is the number of stress drop in the range p to $p + dp$ and $p_1 < p < p_2$ ($p_1 > 0$, where p_2 is limited). We report here that a relationship of this type holds (with specific parameters c_1 and α) for 15 different sets of earthquake occurring in Japan, California, Mexico, Nevada, Italy and the Aleutian arc. We also show that the frequency distribution $N(p)$ provides some information about the frequency distribution of ground acceleration during earthquakes.

Figure 1 verifies that the density distribution of the stress drops is a decreasing function of p , and also that equation (1) is a satisfactorily representation of the frequency distribution of p . The stress drops computed by Archambeau (unpublished data) for the Alaskan-Aleutian region and the Japanese region have been obtained by fitting observed seismograms with synthetic ones obtained from source models of the events that have the observed source parameters. The stress drops of the data of Mikumo and Utsu (personal communications) have been computed by various methods including Brune's formula¹⁵. The value of $-1 + \alpha$ obtained from the analysis of Archambeau's data for Japan is -1.2, that from the data of Mikumo and Utsu is -1.4: considering that there are relatively few data in both sets these two values agree well.

Few authors report the accuracy of their data. However, Peppin and Bufe⁹ computed the seismic moment and the stress drop for 40 earthquakes; they did not find large differences between the stress drops measured from vertical P-wave and horizontal S-wave spectral corner frequencies. They also computed the moment and stress drop using records from several different instruments. The scatter of the data is acceptable, 65% of their corner frequencies have precision better than 10%.

The regression lines of Fig. 1 have been computed using the method of the maximum likelihood assuming $\sigma(\log N)/\sigma(\log p) = 1$. The values of $1 - \alpha$ are affected by a mean standard deviation of at most 10%.

To obtain the density distribution of the r.m.s. acceleration of the ground we also need the density distribution $N(l)$ of the linear dimension of the faults l . This has the form $N(l) = c_2 l^{-\nu}$ (ref. 12) with $\nu = 2b + 1$, where b is given by the results from the Gutenberg-Richter frequency (n) magnitude (M) relation $\log n = a - bM$, b is typical of each seismic region and generally varies from 0.5 to 1.

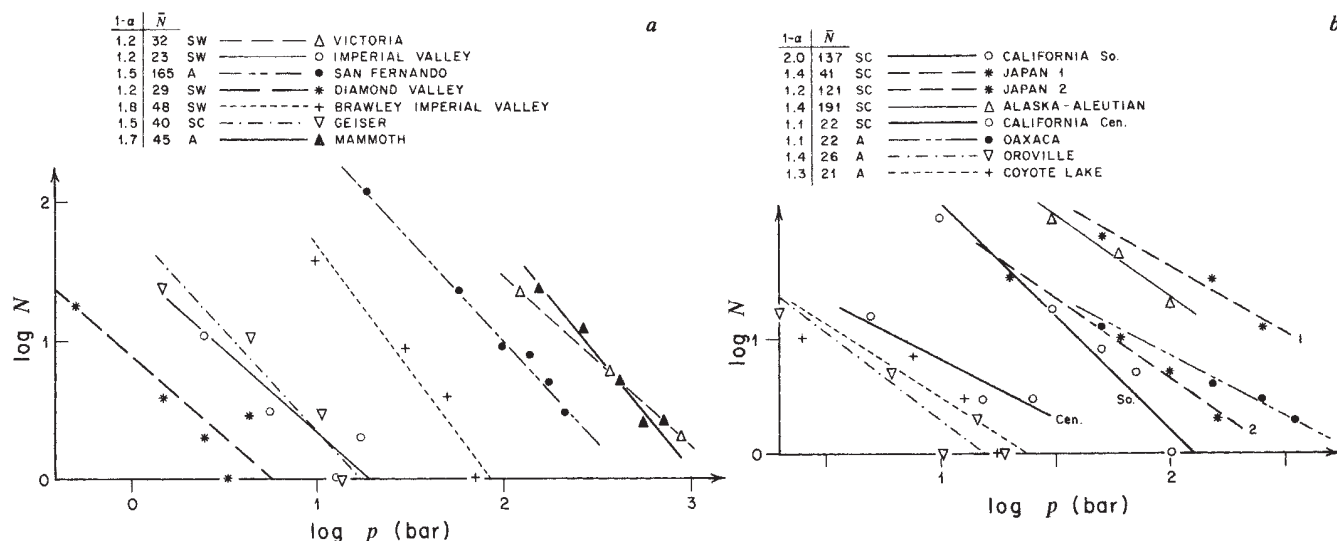


Fig. 1 Linear regression for the stress drops of aftershocks of: *a*, the 1971 San Fernando earthquake¹¹; earthquakes of the Brawley-Imperial Valley 1975 swarm⁵; earthquakes of the Imperial Valley 1977 swarm¹; scattered earthquakes in the Geiser (California) region⁹; earthquakes of the 1978 Victoria (Mexico) swarm⁵; earthquakes of a sequence in Diamond Valley (Sierra Nevada)³; aftershocks of the May 1980 Mammoth earthquake (R. I. Archuleta, personal communication), $\bar{N}$ is the total number of events used in the analysis. SC, Scattered events; A, aftershocks; SW, swarm. *b*, The 1979 Coyote Lake (California) earthquake⁶; aftershocks of the 1978 Oroville (California) earthquake⁴; earthquakes of the 1974 sequence of Central California²; aftershocks of the 1978 Oaxaca (Mexico) earthquake⁷; scattered earthquakes in the Aleutian and Alaskan regions (C. Archambeau, unpublished data); scattered earthquakes in Japan (1) (C. Archambeau, unpublished data); scattered earthquakes in Japan (2) (personal communications from T. Mikumo, and T. Utsu); scattered earthquakes in Southern California¹⁰.

If the density distributions of p and l are known the density distribution of the r.m.s. ground acceleration a , can be calculated using the relationship established by McGuire and Hanks¹⁷

$$a = 0.317 \frac{p}{\rho R} \left(\frac{f_{\max}}{f_0} \right)^{1/2} \quad (2)$$

where ρ is the density, R the hypocentral distance, f_0 the spectral corner frequency of the seismic signal and $f_{\max} = Q\beta/\pi R$ is determined by the distance from the point to the epicentre R , the velocity of the waves β and the specific dissipation Q . Substituting¹⁵

$$f_0 = 2.34\beta/2\pi l$$

in equation (2) we find

$$a = \frac{0.29}{\rho} (Q)^{1/2} \frac{p}{R} \left(\frac{l}{R} \right)^{1/2} = kpl^{1/2}, \quad k = \frac{0.52}{\rho R} \left(\frac{f_{\max}}{\beta} \right)^{1/2} \quad (3)$$

The cumulative distribution function $n(a)$ of a can be determined from the density distributions (1) of l and p as follows. If p_1, p_2, l_1, l_2 are the minimum and maximum values of p and l , of the seismic region¹⁶ then by integrating¹² in the domain of the plane l, p defined by the minimum and maximum values of p and l , we find that the cumulative distribution of a for $a > kp_1\sqrt{l_2}$ (or $a > kp_2\sqrt{l_1}$ whichever is larger) is

$$n(a) = \frac{D}{1-\nu} \left\{ \frac{p_2^{\alpha} l_2^{1-\nu}}{\alpha} - \frac{(2\nu-2)l_2^{1-\nu-\alpha/2} a^{\alpha}}{\alpha(2\nu+\alpha-2)k^{\alpha}} - \frac{p_2^{2\nu+\alpha-2} a^{2-2\nu}}{k^{2-2\nu}(2\nu+\alpha-2)} \right\} \quad (4)$$

In the crust it is reasonable to assume $\rho = 2.8$ and $Q = 300$ for the seismic waves of interest, this gives $k = 1.81R^{-3/2}$. Because $\alpha < 0, \nu > 2$ it is seen that $n(a)$ is an increasing function of p_2 and decreasing function of a and R .

For small values of a in the range $p_1 k \sqrt{l_2}$ to $p_2 k \sqrt{l_1}$ such as those considered here the second term of equation (4) is small relative to the third. Because $\nu \approx 2.8$, $n(a)$ is nearly proportional to $(aR^{3/2})^{-\delta}$ ($\delta \approx 3.5$).

Note that, considering equation (3) and the density distributions of p and l given by equation (1), a given pair of values of a and R , correspond to an infinite number of values of M and that, of all the values of M associated with this pair, Fig. 2, the

minimum values of M is that which has the highest probability of occurrence. This implies that the return periods of the accelerations of the ground caused by earthquakes are much longer than previously estimated through empirical relations between M and a .

If we take into account that the seismicity is not concentrated in a point at distance R but covers an area, which we assume here for simplicity to be a sector of a circular ring defined by the angles ϑ_1, ϑ_2 and by the radii R_1 and R_2 , then $n(a)$ is

$$n(a) = \frac{Dp_2^{\alpha} l_2^{1-\nu}}{1-\nu} \left\{ \frac{1}{\alpha} - \frac{4(\nu-1)(a/kp_2\sqrt{l_2})^{\alpha} (R_3^{3\alpha+2} - R_1^{3\alpha+2})}{(\frac{3}{2}\alpha+2)(2\nu+\alpha-2)(R_2^2 - R_1^2)\alpha} - \frac{2(a/kp_2\sqrt{l_2})^{2-2\nu} (R_3^{5-3\nu} - R_1^{5-3\nu})}{(2\nu+\alpha-2)(5-3\nu)(R_2^2 - R_1^2)} \right\}$$

where $R_3 = R_2$ or $(p_2\sqrt{l_2})^{2/3}$ whichever is smaller.

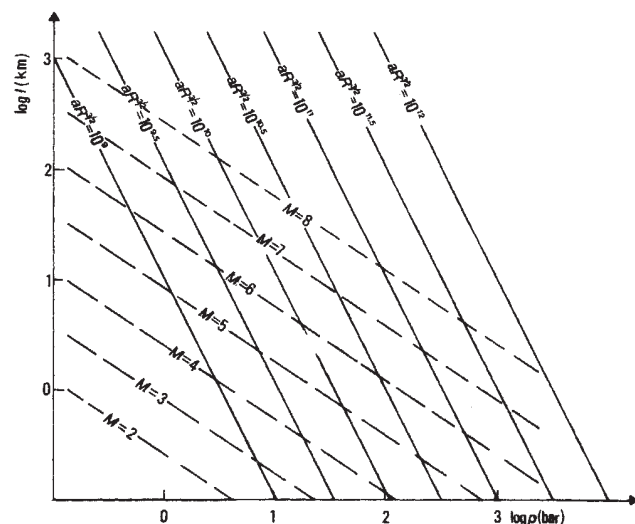


Fig. 2 Isolines of magnitudes and r.m.s. acceleration a in the plane of the linear dimensions of faults l and stress drops p . The lines $aR^2 = \text{constant}$ have been computed from equation (3), the lines $M = \text{constant}$ have been computed according to the formula $10^{11.1+1.46M} = \frac{16}{7}(l^3 p^2 / 2\mu)$ of ref. 13.

In California $\nu = 2.78$, $\alpha = -1$, (ref. 13); a reasonable value of p_2 is 500 bar (refs 10, 13); we may tentatively consider a maximum linear dimension of a fault $l_2 = 5 \times 10^6$. Using equation (15) of ref. 12 with $p_2 = 5 \times 10^8$, $p_1 = 5 \times 10^5$ and the data of Epstein and Lomnitz¹⁸ on Northern California Earthquakes one obtains $D = 10^{16.35}$ (ref. 12). For $a = 50$ Gal, $R_2 = 8R_1 = 8 \times 10^6$, equation (4) gives $n(50) = (17)^{-1}$ or a return period of 17 yr; this is much longer than the return period of $M = 5$ which, as shown by Fig. 2, could cause an acceleration of 50 Gal in the rare case when $l = 100$ m, and $p = 500$ bar.

Applying equation (2) to Southern California¹⁷ and Friuli (E. Faccioli, personal communication) gave r.m.s. accelerations of the ground close to those observed; however, a test on a sequence of earthquakes recently observed in Northern California gave ambiguous results (T. C. Hanks, personal communication). If for some regions a formula for a must be used that is different from equation (2), such as that of Bernreuter¹⁹ $a = (0.19p/\rho R)(l/R)^{0.7}$, the present method would still be valid, and a slightly different form for $n(a)$ would be obtained.

Received 19 November 1980; accepted 25 February 1981.

- Adair, R., Brune, J. N. & Orcutt, J. *Earthquake Not.* **49**, 4 (1978).
- Bakun, W. H., Bufe, G. C. & Stewart, R. M. *Bull. seism. Soc. Am.* **66**, 363 (1976).
- Sommerville, M. R., Peppin, W. A. & Van Wormer, J. R. *Bull. seism. Soc. Am.* **70**, 1547 (1980).
- Fletcher, J. B. *Bull. seism. Soc. Am.* **70**, 735 (1980).
- Hartzell, S. H. & Brune, J. N. *Pageophysics* **115**, 333 (1977).
- Lee, W. H. K., Herd, D. G., Cagnetti, V., Bakun, W. H. & Rapport, A. *U.S. geol. Surv. Open File Rep.* **79**, 1621 (1979).
- Munguia, L. et al. *Geophys. Int.* **17**, 359 (1978).
- Munguia, L. et al. *EOS* **59**, 1131 (1978).
- Peppin, W. A. & Bufe, C. *Bull. seism. Soc. Am.* **70**, 269 (1980).
- Tatcher, A. & Hanks, T. C. *J. geophys. Res.* **78**, 8547 (1973).
- Tucker, B. E. & Brune, J. N. *Geol. Geophys. Stud.* **3**, 65 (1973).
- Caputo, M. *Ann. Geofis.* **4**, 277 (1976).
- Caputo, M. & Console, R. *Tectonophysics* **67**, T13 (1980).
- Caputo, M. *Tectonophysics* **69**, T9 (1980).
- Brune, J. N. *J. geophys. Res.* **75**, 4997 (1970).
- Hanks, T. C. *J. geophys. Res.* **84**, 2235 (1979).
- McGuire, R. K. & Hanks, T. C. *Bull. seism. Soc. Am.* **5**, 70 1910 (1980).
- Epstein, B. & Lomnitz, C. *Nature* **211**, 1197 (1966).
- Bernreuter, D. L. *OECD-NEA/CSNI Rep. No. 28*, 2, 429 (1979).

Can Kimberlites be generated from an ordinary mantle?

J. D. Kramers*, C. B. Smith†, N. P. Lock‡, R. S. Harmon§ & F. R. Boyd||

* Department of Earth Sciences, University of Leeds, Leeds LS2 9JT, UK

† Bernard Price Institute of Geophysical Research, University of The Witwatersrand, Johannesburg, South Africa

‡ Department of Geology, University of Sheffield, Sheffield S10 2TN, UK

§ Scottish Universities Research and Reactor Centre, East Kilbride, Glasgow G75 0QU, UK

|| Geophysical Laboratory, 2801 Upton Street, NW, Washington, DC 20008, USA

Kimberlites are characterized by a high concentration of incompatible elements, including light rare earths. However, Nd isotopic evidence indicates that their source regions do not have a long history of enrichment. If kimberlites can be generated from non-enriched mantle sources by simple partial melting processes, this implies that crystal-liquid partition coefficients for some trace elements are lower for kimberlitic liquids than for basalts or andesites. From a comparison of clinopyroxene megacrysts (regarded as equilibrated with kimberlite at depth) with existing kimberlite data we argue here that low crystal-liquid partition coefficients for the rare earths are plausible.

Experimental work on the melting relations of model peridotite systems charged with CO_2 has shown that liquids of carbonatitic and kimberlitic composition may be produced by a small degree of partial melting of a carbonated peridotite^{1,2}. This

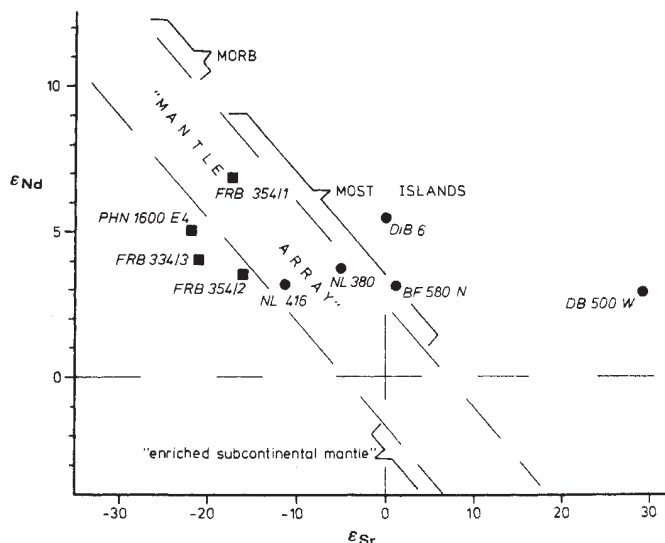


Fig. 1 ϵ_{Nd} plotted against ϵ_{Sr} showing kimberlite (●) and diopside megacryst (■) samples. The age³⁷ used for correction is 90 Myr. The 'mantle array' contains data from mid-ocean ridge basalts (MORB) and most oceanic islands as well as 'enriched subcontinental mantle' as exemplified by diopsides from peridotite nodules in kimberlite⁹.

$$\epsilon_{\text{Nd}} = \left[\frac{(^{143}\text{Nd}/^{144}\text{Nd})_{\text{init}} \text{ of sample}}{(^{143}\text{Nd}/^{144}\text{Nd})_{\text{of CHUR}}} - 1 \right] \times 10^4$$

where t denotes the age of the sample. Analogous for ϵ_{Sr} . $(^{147}\text{Sm}/^{144}\text{Nd})_{\text{CHUR}} = 0.194$; $(^{143}\text{Nd}/^{144}\text{Nd})_{\text{CHUR}}(\text{present day}) = 0.51264$; $(^{87}\text{Rb}/^{86}\text{Sr})_{\text{UR}} = 0.0827$; $(^{87}\text{Sr}/^{86}\text{Sr})_{\text{UR}}(\text{present day}) = 0.7045$.

melting is likely to be caused by a change in the P - T conditions rather than an incursion of CO_2 , because CO_2 cannot exist as a free fluid phase in a silicate mantle at pressures higher than ~ 32 kbar, but is bound by the formation of carbonate minerals². The melting constitutes a single process and, therefore, trace-element enrichment in the liquid relative to its source is limited by solid-liquid partition coefficients.

An assessment of the time-integrated enrichment or depletion in incompatible elements, including light rare earths (LREs), in the sources of volcanic rocks can be made from Nd and Sr isotope data. $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of fresh kimberlites are ~ 0.704 (ref. 3), indicating time integrated source Rb/Sr ratios lower than that of the 'uniform reservoir' (UR)⁴ also known as 'bulk earth'⁵. Recent Sm-Nd work on kimberlites ranging in age between 1,300 and 90 Myr (ref. 6) seems to point to time-integrated source Sm/Nd ratios close to that of the 'chondritic uniform reservoir' (CHUR)⁴ or 'bulk earth'⁵. In contrast, combined Sr and Nd work on American kimberlites⁷ has suggested both time integrated large-ion lithophile (LIL) and LRE element enrichment and depletion for their sources. We have measured the $^{143}\text{Nd}/^{144}\text{Nd}$ and $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of five Cretaceous Southern African kimberlites, using methods described elsewhere⁸, and the results are listed in Table 1. In the $\epsilon_{\text{Nd}}/\epsilon_{\text{Sr}}$ plot⁴ (Fig. 1) the two dyke kimberlites from Letseng la Terai (NL 380 and 416) and the Bultfontein sample (BF 580 N) plot close to the 'mantle array' which is defined by both oceanic and continental volcanics⁵ as well as some mantle derived xenoliths⁹. They indicate a slight time-integrated LIL and LRE element depletion of their sources. Kimberlite samples DiB 6 and DB 500 W plot well to the right of the mantle array, which is thought to be the result of Sr exchange with circulating groundwater³. Any effect on the $^{143}\text{Nd}/^{144}\text{Nd}$ ratios is likely to have been a lowering one⁴; thus these two kimberlites were probably also derived from sources with a time integrated LREE depletion. From published Sr and Nd isotopic data on kimberlites^{3,6,7} and our results, the source regions of kimberlites seem to be similar in their LIL and LRE element characteristics to the sources of ocean island volcanics. These are characterized

Table 1 Isotopic data on kimberlites and clinopyroxene megacrysts

Sample	$^{87}\text{Sr}/^{86}\text{Sr}$	$^{143}\text{Nd}/^{144}\text{Nd}$	$^{87}\text{Rb}/^{86}\text{Sr}$	$^{147}\text{Sm}/^{144}\text{Nd}$	$\delta^{18}\text{O}$ (SMOW)
Kimberlites					
NL 380 (Dyke, Letseng la Terai, Lesotho)	0.70439 $\pm$ 3	0.51278 $\pm$ 4	0.281	0.0936	
NL 416 (Dyke, Letseng la Terai, Lesotho)	0.70386 $\pm$ 2	0.51274 $\pm$ 3	0.218	0.0916	
DiB 6 (Wesselt Mine, Kimberley, South Africa)	0.70456 $\pm$ 6	0.51285 $\pm$ 2	0.127	0.0717	
BF 580 N (Bultfontein Pipe, Kimberley, South Africa)	0.70461 $\pm$ 3	0.51274 $\pm$ 2	0.119	0.09	
DB 500 W (De Beers Pipe, Kimberley, South Africa)	0.70691 $\pm$ 10	0.51273 $\pm$ 2	0.353	0.915	
Megacrysts					
FRB 334/3 (Camutue Pipe, Angola)	0.70293 $\pm$ 20	0.51283 $\pm$ 3	0.0082	0.147	+6.57
PHN 1600 E4 (Thaba Putsoa Pipe, Lesotho)	0.70280 $\pm$ 20	0.51289 $\pm$ 3	0.0023	0.168	+6.56
FRB 354/1 (Kimberley Pool, Kimberley, South Africa)	0.70317 $\pm$ 18	0.51302 $\pm$ 4	0.0026	0.191	+6.35
FRB 354/2 (Kimberley Pool, Kimberley, South Africa)	0.70326 $\pm$ 14	0.51281 $\pm$ 2	0.004	0.168	
FRB 354/3 (Kimberley Pool, Kimberley, South Africa)					+6.88
FRB 354/4 (Kimberley Pool, Kimberley, South Africa)					+6.50
IM 7-C1 (Front Range kimberlites, Wyoming)	0.70360 $\pm$ 10		0.003		
Standards					
NBS 987 Sr (66 runs since 1977)	0.710272 $\pm$ 38 (1 s.d. of population)				
La Jolla Nd standard (45 runs since 1977)	0.511880 $\pm$ 36 (1 s.d. of population)				

Errors quoted on Sr and Nd isotope ratios are 2 standard deviations of the mean (except for standards). Errors on Rb/Sr and Sm/Nd ratios estimated at 2%. $^{143}\text{Nd}/^{144}\text{Nd}$ ratios normalized to $^{146}\text{Nd}/^{144}\text{Nd} = 0.7219$.

by trace element and isotopic heterogeneity, varying from no differentiation (relative to CHUR) to moderate LIL and LRE element depletion⁵. They also exhibit the same heterogeneous, mostly radiogenic lead isotopic signature as kimberlites¹⁰.

Geochemical work on kimberlites¹¹⁻¹³ has highlighted their strong enrichment in incompatible elements, including LREE. One explanation invoked for this is that their mantle sources were pre-enriched^{13,14}, but long-term enrichment of kimberlite sources can be discounted on the grounds of the isotopic evidence discussed above, with the possible exception of highly micaceous kimberlites, which can have genuinely high initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratios³. From isotopic^{9,10,15} and trace element¹⁶⁻¹⁸ studies on mantle derived xenoliths it has been shown that provinces exist in the upper mantle underneath Southern Africa which are anomalously enriched in LIL and LRE elements and have been so for a long period of geological time⁹. However, these were found to occur at shallower levels than the source regions of kimberlites^{19,20} and their high $^{87}\text{Sr}/^{86}\text{Sr}$ ratios preclude a genetic link with most kimberlites.

The idea that kimberlites, with their strong enrichment in incompatible elements, could be derived from an 'ordinary

mantle' which was slightly depleted in these elements, is an apparent paradox. Mechanisms which could resolve this paradox include:

(1) Transport of some incompatible elements as complexes in a CO_2 or H_2O vapour phase¹¹. This mechanism may account for the great heterogeneities found in diatremes at near-surface level, but is unlikely to operate at the depth of origin of kimberlites^{1,2}.

(2) Metasomatic enrichment of the kimberlite source region shortly before melting, as proposed for some alkali basalts²¹. There is no evidence excluding this process, but for kimberlites the metasomatizing agent would have to be a melt^{1,2} and the mechanism therefore involves 'two-stage' melting.

(3) Simple melting with trace-element partitioning governed by lower solid-liquid partition coefficients than are found in basaltic systems. This alternative is examined below, using K, Rb, Sr and REE data on subcalic clinopyroxene megacrysts.

Among the otherwise variable inclusion population in kimberlite pipes, megacrysts (monomineralic inclusions over ~5 mm in size) of subcalic clinopyroxene, garnet, olivine, orthopyroxene, phlogopite and ilmenite stand out by being almost ubiquitous. Subcalic clinopyroxene megacrysts typically indicate equilibration temperatures^{19,22,23} around 1,300–1,400 °C. Chemically they do not closely resemble clinopyroxenes in xenoliths (although some similarities exist with Northern Lesotho 'sheared ilherzolites'^{17,19}), and unlike other types of inclusions in kimberlites they consistently have low $^{87}\text{Sr}/^{86}\text{Sr}$ ratios¹⁵. They have been regarded as phenocrysts crystallized from the kimberlite magma at depth^{22,24}, which is particularly plausible for the low-Cr (0.3–0.6% Cr_2O_3) variety²². Because their composition usually indicates a whole range of temperatures of equilibration^{19,22}, this interpretation implies a diapiric origin for kimberlite²⁵, with fractional crystallization of phenocrysts occurring over a considerable depth range²².

To test the isotopic consistency of the phenocryst hypothesis, we have measured $^{143}\text{Nd}/^{144}\text{Nd}$, $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{18}\text{O}/^{16}\text{O}$ ratios of four subcalic low-Cr diopside megacrysts from Southern Africa and the $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of a similar diopside megacryst from the Colorado-Wyoming Front Range kimberlite province. Descriptions and major element analyses of the samples have been published elsewhere^{19,22,23,26}. The isotopic results are listed in Table 1. The ϵ_{Nd} and ϵ_{Sr} values (Fig. 1) plot close to the mantle

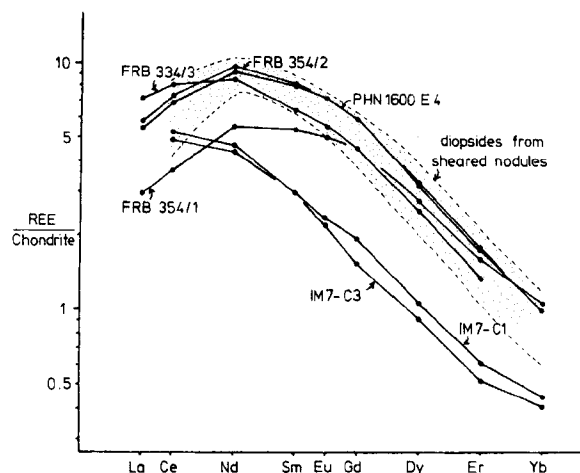


Fig. 2 Chondrite normalized rare earth abundances in diopside megacrysts. Stippled area, field for diopsides from Northern Lesotho sheared ilherzolite nodules analysed by Shimizu¹⁷.

array in a 'moderately depleted' region. We have argued above that the kimberlite points in Fig. 1 were probably moved to the right (and possibly slightly downward) by interaction with groundwater, and might originally have plotted close to the mantle array in a similarly depleted region. The $\delta^{18}\text{O}$ values of the megacrysts (Table 1) are uniform around +6.5, a value higher than that obtained for most ultramafic xenoliths, but similar to phlogopite megacrysts in kimberlite²⁷. Lead isotope ratios of the four Southern African clinopyroxene megacrysts (J.D.K., unpublished data) plot within the field defined by Southern African Cretaceous kimberlites¹⁰.

Although the isotopic data discussed above do not prove that subcalcic clinopyroxene megacrysts crystallized from the kimberlite magma at depth, they do not conflict with this hypothesis, which has also stood up to petrological scrutiny²². Below, assuming equilibrium as a working hypothesis, we derive clinopyroxene–kimberlitic liquid partition coefficients for K, Rb, Sr and the REE and compare them with literature values for basaltic systems. Rb, K, Sr and REE concentrations measured by isotope dilution are listed in Table 2, including results on the kimberlite and megacryst samples listed in Table 1 and one additional diopside megacryst from the Front Range kimberlites. The chondrite normalized REE patterns of the clinopyroxenes (Fig. 2) are similar to patterns found by Shimizu¹⁷ for clinopyroxenes from Northern Lesotho 'sheared' lherzolite nodules, for which an approximately chondritic whole rock REE pattern was reconstructed from the mineral data¹⁷. The subcalcic clinopyroxene megacrysts could thus be in equilibrium with a non-enriched mantle, either directly or through a liquid.

The great heterogeneity of kimberlite occurrences and the variation between them means that the approach to the derivation of partition coefficients can only be semiquantitative, and we use an 'average kimberlite'. Its REE pattern is derived from the mean values of 23 kimberlite analyses, including the five listed in Table 2 and results from the literature^{11–13,28}, and its K, Rb and Sr concentrations are taken from values in Table 2 together with 11 analyses from the literature^{10,15}, avoiding upper diatreme facies kimberlites which can be heavily depleted in alkalis. The values for this 'average kimberlite' are also listed in Table 2, along with standard deviations. The apparent clinopyroxene–kimberlitic liquid partition coefficients listed in Table

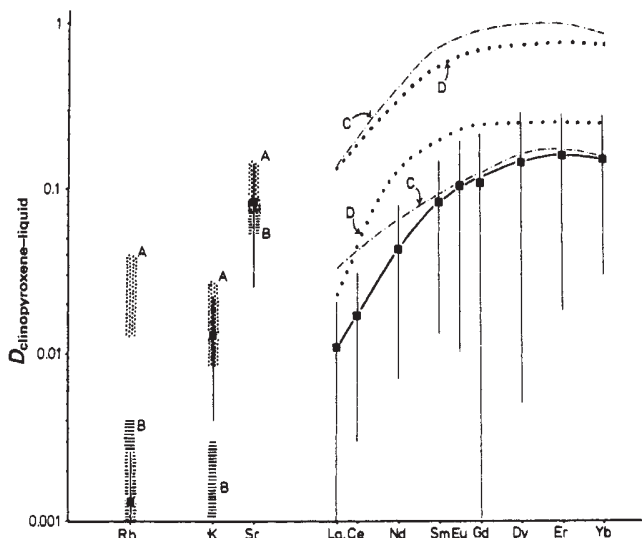


Fig. 3 Apparent clinopyroxene–kimberlitic liquid partition coefficients found in the present investigation (■, with one standard deviation error bars) compared with ranges of literature values for basaltic liquids. A, Philpotts and Schnetzler³³; B, Shimizu²⁹; C, boundaries of field for data of Onuma *et al.*³² and Schnetzler and Philpotts³⁴; D, boundaries of field for data of Nicholls and Harris³¹ and work reviewed by Irving³⁰.

2 have been derived from the average of the clinopyroxene analyses and the 'average kimberlite'. In Fig. 3 these apparent partition coefficients are compared with results obtained from experimental^{29–31} and natural^{32–34} systems. The value for Sr agrees well with the data from clinopyroxene–basaltic liquid studies^{29,33}. For K and Rb there is a considerable discrepancy between Shimizu's²⁹ experimental results and Philpotts' and Schnetzler's³³ data (Fig. 3). Our apparent partition coefficient for K agrees with the latter, whereas our value for Rb conforms to Shimizu's results. The apparent REE clinopyroxene–kimberlitic liquid partition coefficients follow a pattern which is strikingly similar in shape to those derived from both natural

Table 2 Trace element concentrations (p.p.m.) and apparent partition coefficients clinopyroxene–kimberlitic liquid

Sample	Rb	K	Sr	La	Ce	Nd	Sm	Eu	Gd	Dy	Er	Yb
Kimberlites												
NL 379 (Letseng la Terai)	64	4441	932	79.6	154	62.0	9.83	2.65	6.61	3.50	1.21	0.74
NL 380	95	7972	977	168	324	125	19.3	4.95	12.4	6.77	2.53	1.91
NL 416	63	5894	836		193	74.7	11.3	2.88	7.04	3.51	1.23	0.75
DiB 6	120	16700	2727	258	465	190	22.6	6.92	20.5	9.30	2.90	1.62
DB 500 W	72		606	65.1	129	57.4	8.71	1.86	5.68	3.20	1.30	1.00
Megacrysts												
FRB 334/3	0.31	237	112	2.35	7.02	5.35	1.30	0.42	1.22	0.85	0.31	
PHN 1600 E4	0.07	184	89	1.77	6.00	5.80	1.61	0.54	1.60	1.08	0.40	0.22
FRB 354/1	0.06	173	70	0.98	3.16	3.45	1.09	0.38	1.22	0.93	0.37	0.23
FRB 354/2	0.12	228	89	1.89	6.24	6.00	1.67	0.55	1.64	1.10	0.41	0.23
IM 7-C1	0.15	249	126		4.45	2.90	0.60	0.18	0.53	0.36	0.14	0.10
IM 7-C3 (Wyoming)		332			4.21	2.75	0.60	0.17	0.41	0.31	0.12	0.09
average	0.14 ±0.10	233 ±57	97 ±22	1.75 ±0.57	5.18 ±1.46	4.37 ±1.50	1.15 ±0.47	0.37 ±0.17	1.10 ±0.52	0.77 ±0.35	0.29 ±0.13	0.18 ±0.07
Average kimberlite (see text)												
	108 ±35	17600 ±8000	1175 ±550	165 ±92	310 ±160	102 ±50	14 ±6	3.6 ±1.6	10.4 ±6	5.3 ±2.7	1.83 ±0.8	1.20 ±0.5
Apparent partition coefficient clinopyroxene–kimberlite												
	0.0013 ±0.0013	0.013 ±0.009	0.083 ±0.058	0.011 ±0.010	0.017 ±0.014	0.043 ±0.036	0.082 ±0.069	0.103 ±0.093	0.106 ±0.111	0.145 ±0.140	0.158 ±0.140	0.150 ±0.121

Errors listed with averages are standard deviations of populations. Errors quoted for partition coefficients were calculated from these standard deviations.

basaltic systems^{32,34} and experimental work^{30,31} (Fig. 3), but the values are up to an order of magnitude lower. A general observation in studies on both synthetic and natural systems^{30,31,34} has been that crystal-liquid partition coefficients decrease with decreasing SiO₂ and increasing MgO content (that is, decreasing polymerization) of the liquid, and kimberlites are much poorer in SiO₂ (30–40%) and richer in MgO (generally 20–30%) than any siliceous volcanic rocks used in partition coefficient studies. A decrease in polymerization of the melt and thus in partition coefficient values could also result from dissolved H₂O (ref. 31). Although dissolved CO₂ may not have a depolymerizing effect, it might lower REE partition coefficients by complexing with REE, as is in evidence in vapour-phase CO₂ (ref. 35). Finally, crystal-liquid partition coefficients generally decrease with increasing temperature^{30,31}. Without attempting a quantitative assessment of these effects, it can be predicted from the physicochemical differences between a kimberlitic and a basaltic liquid in their respective source regions that REE crystal-liquid partition coefficients would be considerably lower for the former. The low partition coefficients found for the REE in this work would thus not be unreasonable. Partition coefficients for Sr seem to be less affected by the basicity of the melt than those for the REE³³, and the present results are consistent with this observation.

Because the low apparent values found for clinopyroxene-kimberlitic liquid REE partition coefficients, compared with basaltic systems, would be largely understandable as resulting from properties of the liquid rather than the solid, the REE partition coefficients for other minerals in equilibrium with a kimberlitic liquid could be expected to be correspondingly lowered. A kimberlite with a typical strongly LREE-enriched pattern could thus be in REE equilibrium with ordinary mantle material such as forms the source of ocean island basalts. However, equilibrium with the original parent material is only a boundary condition of any melting model and implies either an almost infinitely small percentage of partial melt, or complete equilibration of a relatively small amount of melt with a very large volume of rock similar to the source rock. This ultimately happens in zone refining³⁶, but could also be achieved if the melt existed as intergranular films in a source rock undergoing penetrative deformation, for instance, if melting occurred in a rising diapir²⁵.

Although the above argument is in some ways speculative, the hypothesis of kimberlite generation from an ordinary mantle by a single melting process merits consideration from a trace element point of view.

We thank J. C. Roddick for analytical assistance and W. Wilkinson for mineral purification. Isotope research at Leeds is supported by the NERC which is thanked by J.D.K. for a grant. P. J. Betton, R. A. Cliff, M. J. Norry and M. F. Thirlwall provided constructive criticism.

Received 17 November 1980; accepted 17 February 1981.

1. Eggler, D. H. *Yb Carnegie Instn Wash.* **74**, 468–474 (1975).
2. Wyllie, P. J. *J. Geol.* **86**, 687–713 (1978).
3. Barrett, D. R. & Berg, G. W. *Phys. chem. Earth* **9**, 619–636 (1975).
4. DePaolo, D. J. & Wasserburg, G. J. *Geochim. cosmochim. Acta* **43**, 615–628 (1979).
5. O'Nions, R. K., Carter, S. R., Evensen, N. M. & Hamilton, P. J. A. *Rev. Earth & planet. Sci.* **7**, 11–38 (1979).
6. Basu, A. H. & Tatsumoto, M. *Science* **205**, 398–401 (1979).
7. DePaolo, D. J. *U.S. Geol. Surv. Open File Rep.* 1978–701, 81–93 (1978).
8. Hawkesworth, C. J. *et al. Earth planet Sci. Lett.* **42**, 45–57 (1979).
9. Menzies, M. & Murthy, V. R. *Nature* **283**, 634–636 (1980).
10. Kramers, J. D. *Earth planet. Sci. Lett.* **34**, 419–431 (1977).
11. Fesq, H. W., Kable, E. J. D. & Gurney, J. J. *Phys. chem. Earth* **9**, 687–707 (1975).
12. Mitchell, R. H. & Brunfelt, A. O. *Phys. chem. Earth* **9**, 671–686 (1975).
13. Frey, F. A., Ferguson, J. & Chappell, B. W. *2nd Int. Kimberlite Conf. Ext. Abstr.* (Carnegie Institution, Washington DC, 1977).
14. Kramers, J. D. *Earth planet. Sci. Lett.* **42**, 58–70 (1979).
15. Barrett, D. R. *Phys. chem. Earth* **9**, 637–654 (1975).
16. Shimizu, N. *Phys. chem. Earth* **9**, 655–669 (1975).
17. Shimizu, N. *Earth planet. Sci. Lett.* **25**, 26–32 (1975).
18. Nixon, P. H., Rogers, N. W., Gibson, I. L. & Grey, A. A. *Rev. Earth planet. Sci.* (in the press).
19. Boyd, F. R. & Nixon, P. H. in *Lesotho Kimberlites* (ed. Nixon, P. H.) 67–75 (Lesotho National Development Corporation, 1973).
20. Dawson, J. B., Gurney, J. J. & Lawless, P. J. *Nature* **257**, 299–300 (1975).
21. Menzies, M. & Murthy, V. R. *Earth planet. Sci. Lett.* **46**, 323–334 (1980).

22. Eggler, D. H., McCallum, M. E. & Smith, C. B. in *The Mantle Sample*, 213–226 (American Geophysical Union, 1979).
23. Boyd, F. R. & Nixon, P. H. *Geochim. cosmochim. Acta* **42**, 1367–1382 (1978).
24. Nixon, P. H., von Knorring, O. & Rooke, J. M. *Am. Miner.* **48**, 1090–1132 (1963).
25. Green, H. W. II & Gueguen, Y. *Nature* **249**, 617–620 (1974).
26. Boyd, F. R. & Danchin, R. V. *Am. J. Sci.* **280-A** (Jackson Vol.) 528–549 (1980).
27. Sheppard, S. M. F. & Dawson, J. B. *Phys. chem. Earth* **9**, 747–763 (1975).
28. Smith, C. B. thesis, Colorado State Univ., Fort Collins (1977).
29. Shimizu, N. *Geochim. cosmochim. Acta* **38**, 1789–1798 (1974).
30. Irving, A. J. *Geochim. cosmochim. Acta* **42**, 743–770 (1978).
31. Nicholls, A. & Harris, K. L. *Geochim. cosmochim. Acta* **44**, 287–308 (1980).
32. Onuma, N., Higuchi, H., Wakita, H. & Nagasawa, H. *Earth planet Sci. Lett.* **5**, 47–51 (1968).
33. Philpotts, J. A. & Schnetzler, C. *Geochim. cosmochim. Acta* **34**, 307–322 (1970).
34. Schnetzler, C. & Philpotts, J. A. *Geochim. cosmochim. Acta* **34**, 331–340 (1970).
35. Wendlandt, R. F. & Harrison, W. J. *Contr. Miner. Petrol.* **69**, 409–419 (1979).
36. Harris, P. G. & Middlemost, E. A. K. *Lithos* **3**, 77–88 (1969).
37. Davis, G. L. *2nd Int. Kimberlite Conf. Ext. Abstr.* (Carnegie Institution, Washington DC, 1977).

Late Phanerozoic extent of dry land

J. Graham Cogley

Department of Geography, Trent University, Peterborough, Ontario, Canada K9J 7B8

The extent of dry land during post-Carboniferous time seems to have fluctuated about a constant mean—no secular decrease in marine inundation of the continents can be identified. Such a decrease was first proposed by Egged¹, who tried to show, with palaeogeographical measurements, that the seas have withdrawn steadily during the Phanerozoic; he deduced on this basis that the Earth is expanding. The idea of an expanding Earth has had little support², but the steady marine recession seems to be more widely accepted. Hallam³, for example, confirmed Egged's finding, and explained it as arising from orogenesis. My new estimate of the extent of the dry land, however, casts doubt on these earlier results, largely by taking into account continental drift and the progressive extension of shelf seas around the young Atlantic and Indian Oceans.

Wise⁴ has argued that the secular withdrawal arises from the limitations inherent in palaeogeographical maps. An inevitable loss of resolution at more remote epochs imposes a temporal bias of uncertain direction, and a spatial bias towards shorter, simpler shorelines. These biases can be eliminated by sampling uniformly long intervals and by simplifying more recent shorelines; the resulting estimates are uniformly very inaccurate. Furthermore, the spectrum of sea-level change has a marked peak corresponding to periods of ~1–2 Myr (ref. 5). Such periods cannot be resolved easily (and perhaps not at all) by palaeogeographical methods, and they are much shorter than the present limits of palaeomagnetic resolution. Thus any palaeogeographical map will probably include information from more than one transgressive cycle, and in particular from the possibly asynchronous transgressive maxima of different regions. An overestimate of the extent of the seas will appear in any map which samples a period longer than ~1–2 Myr, and if the sample period grows shorter with time the extent of the seas will apparently decrease with time.

Notwithstanding these observations, Hallam's³ data suggest secular withdrawal of the seas from the USSR and North America, and he concludes that the trend is real, and due most probably to orogenic thickening of the continents combined with variation in the length of oceanic ridges. It is not clear whether there ought to be a trend in ridge length (although large fluctuations are possible), but there may be a secular decrease in ridge volume, with a sea-level equivalent as great as –80 m (ref. 6) since –500 Myr if ridge volume is proportional to a 20% decrease in internal heat production; this latter estimate seems quite possible^{7,8}, although an 8–10% decrease may be more realistic. There are compelling arguments for a secular increase in the mass of the continents, but they are consistent with

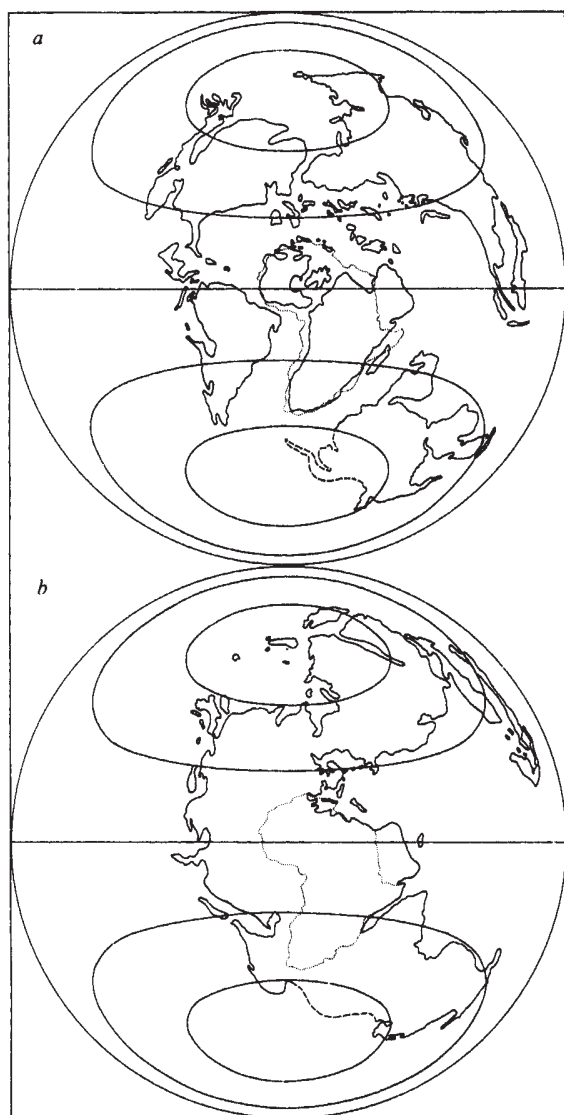


Fig. 1 Palaeogeographical reconstructions for a, -100 Myr and b, -200 Myr. Lambert equal-area projection of the whole sphere; parallels every 30°.

constant thickness (accommodating the added mass by an increase in area), and most of the increase may have been complete long before the Phanerozoic⁹. The continents may have thickened by net horizontal shrinkage without an increment of mass, as perhaps in the Appalachians¹⁰, where wide, thick wedges of sediment and basement were thrust horizontally and stacked on top of each other in the Palaeozoic. Orogenic thickening, however, may be counterbalanced by the thinning recently described^{11,12} as a precursor to the growth of sedimentary basins.

Fluctuations of sea level, as opposed to trends, may appear in history due to several causes¹³⁻¹⁵. Southam and Hay have shown that changes since -120 Myr can be explained by glaciation, episodic fast spreading at the ridges and a decrease in total ocean basin volume. Harrison *et al.* consider ocean sedimentation in more detail, and estimate the influence on apparent sea level of changes in the area-altitude distribution of the continents. Bond has also examined the latter influence, from a different perspective. Whatever the mechanisms that alter the sea level, I show here that the palaeogeographical evidence which prompted some of the earlier proposals is doubtful.

Table 1 Estimates of marine withdrawal from the continents

Sample	Withdrawal rate ($10^3 \text{ km}^2 \text{ Myr}^{-1}$)	
	$-270 \leq T \leq 0$	$-582 \leq T \leq 0$
Cog	-6 ± 32	—
Bar	1 ± 43	—
Ter	67 ± 27	48 ± 9
TerH	67 ± 27	63 ± 9
Str	188 ± 106	101 ± 23
Hal1	55 ± 39	134 ± 18
Hal2	131 ± 100	119 ± 26
Wis	21 ± 29	15 ± 9

Sources: Cog, this work; Bar, ref. 30; Ter, ref. 1 (GPA), Table 2; TerH, Ter with $12 \times 10^6 \text{ km}^2$ subtracted from five earliest points³³; Str, ref. 1 (GPA), Table 1; Hal1, ref. 3 (1977), Fig. 1 (USSR); Hal2, ref. 3 (1977), Fig. 2 (North America revised); Wis, ref. 4, Fig. 3 (North America: this work was an attempt to show that the rates of Str and TerH could be made to approach zero by sampling the record more densely).

There is no reason to suppose *a priori* that the continental shelves, which are below sea level today, have histories of inundation which differ significantly from those of areas which are above sea level today. It is not known whether at any instant the periphery of a continent is more likely to be inundated than its interior. Historically, however, a pattern of old fissures and sutures criss-crossing the continental interiors reveals that regions now far from oceans were once peripheral to old continents; conversely, all the modern Atlantic and Indian shelves were in continental interiors ~200 Myr ago. Earlier palaeogeographical analyses have ignored modern shelves for lack of offshore data, but consequently large areas of dry land have probably been left out. Indeed, if only present land areas are considered, then, unless the ancient and modern shorelines coincide everywhere, estimates of the extent of ancient seas must exceed the extent of modern seas. This property of the estimates is an artefact, and is neither logically nor geologically necessary.

Even without data from boreholes or seismic exploration, reasonable judgment can yield histories for modern shelves which are much simpler than those which would have to be contrived, often in conflict with palaeobiogeography, to ensure that they have always remained below sea level. Note that at -200 Myr in Fig. 1 the shelf off eastern North America is above sea level and well away from the shores of Pangaea; the alternative is a narrow seaway connecting Mexico with Scotland.

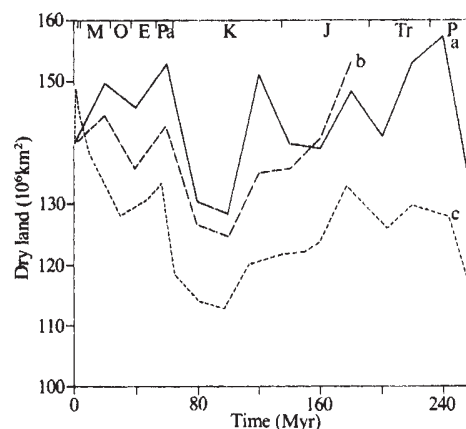


Fig. 2 Extent of dry land estimated from palaeogeographical maps: a, this work; b, ref. 30; c, ref. 1 after ref. 25 (Ter of Table 1). The points at -240 and -260 Myr in a have the same base map³⁴, dated at 250 ± 25 Myr; they are more uncertain, principally as to timing, than the others.

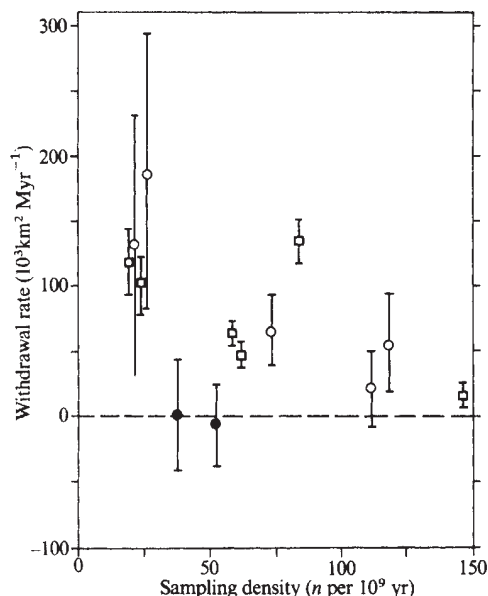


Fig. 3 Estimates of marine withdrawal from the continents. Symbols represent estimates of b (bounded by standard errors) from linear regressions $E = a + bT$, where T is time (adjusted from the Geological Society of London time scale and sources given in ref. 5) and E the extent of dry land (adjusted where necessary to global values). Modern E is set to $140.0 \times 10^6 \text{ km}^2$ to simulate return of polar ice to the oceans³⁵. (Rates are greater, though not significantly, if modern E is set to the observed $148.9 \times 10^6 \text{ km}^2$.) ●, This work and ref. 30; ○, Table 1, column 1; □, Table 1, column 2 (note that the circles and squares are not from independent samples).

Without data from the shelf itself, the reconstruction as adopted is surely the more plausible. In fact, field data^{16,17} show that the late Triassic is either absent, or represented by non-marine sediments penecontemporaneous with graben which suggest East Africa-style rifting. The narrow seaway can be ruled out. Field evidence also serves, with greater or lesser completeness, to rule out seaways between the various once-contiguous fragments of Gondwanaland¹⁸⁻²³.

Note too that the reconstructions of Fig. 1 allow for all known marine sediments within the present shoreline, and that they are conservative where data are lacking and no reasonable judgement can be made. There is, for example, no warrant for a landmass on the shelf south of the Falkland Islands at ~100 Myr.

Figure 1 shows only specimens from a series constructed from 'palaeocontinental' maps²⁴ with shoreline information^{25,26} superimposed. Offshore details were compiled from studies such as those in refs 16-23. These maps were planimetric to yield estimates of the extent of dry land at ~20-Myr intervals from 0 Myr to ~260 Myr (Fig. 2a). The estimates are deficient in several ways. The continental reassemblies²⁴ are not unanimously accepted and are subject to errors of about 5-10° in positioning (95% confidence limits about inferred palaeopoles) and a few millions of years in time. The shoreline data were compiled 30 yr ago²⁵, and are from maps offset by up to 10 Myr from the palaeocontinental maps, but they do have the advantages of internal consistency and global completeness; more recent studies²⁷⁻²⁹ tend to suggest revisions of detail, rather than of broad outline, to the work done by the Termiers. Offshore geology will always be less accessible than onshore, but its uncertainty, in the light of current knowledge, is no longer great enough to justify ignoring it. Percentage errors in offshore estimates can be appreciably worse than onshore errors and still contribute less to total error, because the offshore regions which might once have been land—up to $\sim 50 \times 10^6 \text{ km}^2$ —are less extensive than the onshore regions. About $10 \times 10^6 \text{ km}^2$ of subglacial Antarctica, of which very little is known²⁶, are assumed here to have been above water since ~260 Myr. It is

difficult to place limits on errors from such diverse sources, but an estimate which is, at least partly, independent can be taken from recent work by Barron *et al.*³⁰ (Fig. 2b).

The trends (Table 1, Fig. 3) of Fig. 2a, b are statistically indistinguishable from each other and from zero. They should be compared with the trend seen in the curve of Egyed¹ (Fig. 2c), which was derived from ref. 25 but with data only from modern areas of dry land. The Egyed withdrawal rate and those below it in Table 1 all lack data from modern shelves. The rates in column 2 of Table 1 come from the longer span of the entire Phanerozoic, but only one of them significantly exceeds its counterpart in column 1. That is, most of the withdrawal rates are about as well defined in the past 270 Myr as in the past 580 Myr, indicating that it is reasonable to compare the rates which appear in column 1. Figure 2a lacks points in the Palaeozoic mainly because offshore palaeogeography becomes too speculative^{31,32} for detailed mapping. Estimates of flooding within the present shoreline do not suffer this objection. However, as the arrangement of the continents in remoter times becomes less and less like the present one, the extrapolation from onshore results to putative global results becomes increasingly more doubtful. It may, in fact, be not only reasonable but prudent to concentrate on column 1. In any case, the trends of Fig. 2a, b are well below all the others of Table 1, and their zero secular trend appears also in graphs of relative change of Phanerozoic sea level⁵. (It is not certain, however, whether relative sea-level change and area of dry land are simply related¹³.)

Thus, I conclude that the palaeogeographical evidence is not consistent with a secular withdrawal of the seas from the continents in the second half of the Phanerozoic. It is intriguing to speculate on the biases discussed by Wise⁴, to which Fig. 2a is subject just as are other inundation curves. Does the zero trend of Fig. 2a imply that an 'ideal' version, that is, one with a uniform sampling interval $\approx 2 \text{ Myr}$, would show a secular encroachment of the seas?

I thank NSERC (Canada) for support.

Received 24 November 1980; accepted 18 February 1981.

- Egyed, L. *Geofis. pura appl.* **33**, 42-48 (1956); *Nature* **178**, 534 (1956).
- Glikson, A. Y. *Geology* **7**, 449-454 (1979).
- Hallam, A. *Nature* **269**, 769-772 (1977); **232**, 180-182 (1971).
- Wise, D. U. in *The Geology of Continental Margins* (eds Burk, C. A. & Drake, C. L.) 45-58 (Springer, New York, 1974).
- Vail, P. R., Mitchum, R. M. Jr & Thompson, S. III *Am. Ass. petrol. Geol. Mem.* **26**, 83-97 (1977).
- Armstrong, R. L. *Nature* **221**, 1042-1043 (1969).
- Davies, G. F. *J. geophys. Res.* **85**, 2517-2530 (1980).
- Turcotte, D. L. & Burke, K. *Earth planet. Sci. Lett.* **41**, 341-346 (1978).
- O'Nions, R. K. & Pankhurst, R. J. *Earth planet. Sci. Lett.* **38**, 211-236 (1978).
- Cook, F. A. *et al. Geology* **7**, 563-567 (1979).
- McKenzie, D. *Earth planet. Sci. Lett.* **40**, 25-32 (1978).
- Sclater, J. G. & Christie, P. A. F. *J. geophys. Res.* **85**, 3711-3739 (1980).
- Southam, J. R. & Hay, W. W. *J. geophys. Res.* **82**, 3825-3842 (1977).
- Harrison, C. G. A. *et al. EOS* **61**, 1124 (1980).
- Bond, G. *Geology* **6**, 247-250 (1976); *J. Geol.* **86**, 47-65 (1978).
- Schlee, J. S. *Prof. Pap. P1167* (US Geological Survey, Reston, Virginia, 1980).
- Sheridan, R. E. in *The Geology of Continental Margins* (eds Burk, C. A. & Drake, C. L.) 391-407 (Springer, New York, 1974).
- Baker, B. & Miller, J. *Nature* **199**, 346-348 (1963).
- Blant, G. (ed.) *Sedimentary Basins of the African Coasts, Pt 2, South and East Coasts* (Association of African Geology Surveys, Paris, 1973).
- Hayes, D. E. (ed.) *Antarctic Oceanology II: The Australian-New Zealand Sector*, Antarctic Res. Ser. 19 (American Geophysical Union, Washington DC, 1972).
- Nairn, A. E. M. & Stehli, F. G. (eds) *The Ocean Basins and Margins Vol. 1* (Plenum, New York, 1973).
- Theron, J. C. in *Gondwana Geology* (ed. Campbell, K. S. W.) 61-74 (Australian National University Press, Canberra, 1975).
- Veevers, J. J. & Cotterill, D. *Bull. geol. Soc. Am.* **89**, 337-355 (1978).
- Smith, A. G. & Briden, J. C. *Mesozoic and Cenozoic Palaeocontinental Maps* (Cambridge University Press, 1977).
- Termier, H. & Termier, G. *Histoire Géologique de la Biosphère* (Masson, Paris, 1952).
- Elliott, D. H. *Am. J. Sci.* **275A**, 45-56 (1975).
- Petters, S. W. *J. Geol.* **86**, 311-322 (1978).
- Hallam, A. *J. Geol.* **79**, 129-157 (1971).
- Harrington, H. J. *Bull. Am. Ass. petrol. Geol.* **46**, 1773-1814 (1962).
- Barron, E. J., Sloan, J. L. II & Harrison, C. G. A. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **30**, 17-40 (1980).
- Morel, P. & Irving, E. *J. Geol.* **86**, 535-561 (1978).
- Ziegler, A. M. *et al. Tectonophysics* **40**, 13-51 (1977).
- Holmes, A. *Principles of Physical Geology* 2nd edn, 968-970 (Nelson, London, 1965).
- Briden, J. C., Drewry, G. E. & Smith, A. G. *J. Geol.* **82**, 555-574 (1974).
- Nelson, A. B. *Rand Pap. P-5612* (Rand Corporation, Santa Monica, 1976).

Podzol chronosequences on coastal dunes of eastern Australia

C. H. Thompson

CSIRO Division of Soils, Chancellor's Place, St Lucia, Queensland 4067, Australia

Chronosequences of freely drained podzols, formed in extensive dune systems of quartz sands along the east coast of Australia, offer unique opportunities for comparing temporal changes in different climatic zones during the late Quarternary. The largest areas and greatest number of dune systems occur along the subtropical coast near Brisbane. Here parts of six dune systems have not been covered by younger deposits and have been exposed to weathering since deposition. These areas constitute a relative age sequence on which podzols have formed with solum depths varying with age from <1 m to >20 m. Absolute ages for each system have not been determined but differences in erosion and weathering depths and comparisons with dated dunes elsewhere along the coast indicate an age sequence extending back at least to the last interglacial. Similarities in geomorphology and soils enable correlation with dune systems in other areas to the north and south and so provide a greater potential for examining temporal changes in soils and biological systems than that of previous soil chronosequences¹⁻⁵.

Because Australia largely escaped glaciation and has been relatively stable, tectonically, during the Quarternary, vast quantities of sand have accumulated along its coast⁶. Along the east coast siliceous sands dominate and have apparently been derived from the ocean floor during periods of marine transgression^{6,7}, the large accumulations representing intervals of marine and/or aeolian deposition when sea-level returned to similar positions in successive interglacial periods⁸. Two general types of deposits occur: low sandridges and swales parallel to the coast and widely distributed along its length, and multiple systems of transgressive parabolic dunes at 11 localities between 12 and 40° S latitude (Fig. 1). The parabolic dunes are generally oriented to the south-east onshore winds but their shapes reveal both the control of underlying topography, as in the case of nested dunes, and at exposed sites the influence of onshore winds from other quarters, which lead to multiple-lobe dunes.

Podzols and humus podzols have formed on both types of deposits^{1,9-11} but the high transgressive dunes are more conducive to long-term soil formation because they lack waterables to considerable depth. On these porous, freely drained sands, podzol profile development reaches great depths and on the oldest dunes can exceed 20 m.

The largest area and greatest number of the large dune systems occur between 25 and 28° S. Here large areas of several sand-dune systems form North Stradbroke, Moreton and Fraser Islands and the mainland sandmasses at Cooloola and Peregian^{10,12} (Fig. 1).

Studies of sand-dune systems and associated soils at Cooloola reveal eight periods of dune building. Sizeable areas of six of these dune systems have not been covered by successively younger deposits and have thus been subjected to sub-aerial weathering since deposition. The younger systems (1-4, Fig. 2) consist of parabolic dunes; the dunes of each system are clearly defined by their shape and distribution inland from the coast, the younger dunes being progressively smaller and deposited closer to the coast. The older systems (5 and 6, Fig. 2) extend further inland, have been much modified by erosion and have lost any recognizable dune orientation. They may have been more extensive in the past and parts of 6 may have been reexposed by erosion; however, the strong topographical difference between 5 and 6 makes this unlikely.

The sands are mostly quartz grains with usually <2% of heavy minerals (ilmenite, rutile, zircon and monazite) and some feldspar, but rarely any shell fragments. In all dune systems, the sands are well sorted and have modal grain sizes between 180 and 210 μm . However, the grains in the oldest system are more strongly weathered than those in the youngest (M. P. Green, personal communication), similar to those on nearby Fraser Island¹³.

The unweathered sands are pale yellow-brown due to thin sesquioxidic coatings on the quartz grains. As weathering proceeds, the coatings on grains in the upper part of the profile are mobilized and together with humic compounds deposited lower in the profile to form the friable to firm, iron-humus B horizon (Bhir) of freely drained podzols. With increasing age the top of the Bhir is progressively remobilized and deposited deeper in the profile, the A and B horizons become thicker and more intensely developed, and there is a marked increase in thickness of A2 relative to A1 and B horizons¹¹. In this way solum depth increases with age and is only limited by the thickness of freely drained dune sand and erosion rates since deposition. In the investigated sequences, depth of solum varies from <1 m on the youngest dunes to >20 m on the oldest (Fig. 2). All the sands except those of the youngest dune system are >30 m thick which ensures the pedological unity of the weathering column. This has been established by augering through the soil profiles to unweathered sand (systems 1-4) or into the B horizon (systems 5 and 6). If the trailing arms of parabolic dunes are intact, erosion losses are minimal on dune floors near the apex, low on the crests of trailing arms, and high on other slopes; depth of soil development varies accordingly.

Because both A1 and A2 horizons consist of clean quartz sand from which the sesquioxide coatings have been lost¹⁴, the thickness of these horizons (depth to Bhir) may be used as a measure of relative weathering between analogous sites in different systems. Depth to Bhir is <1.2 m in systems 1-3 (Fig. 2), ~5 m in system 4 and commonly ~11 m in system 5 and 15 m in system 6. However, both systems 5 and 6 have been

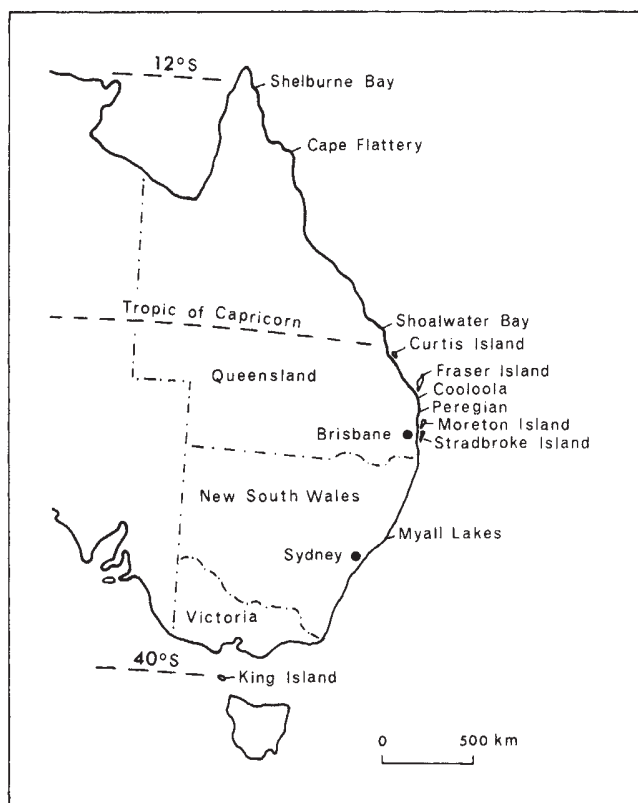


Fig. 1 Locations of main occurrences of parabolic sand-dune systems of quartz sands along the east coast of Australia.

Table 1 Ridge crest profiles from each of the dune systems showing changes in organic carbon and total potassium and phosphorus with depth calculated for 1-m intervals

Depth (m)	Dune systems																	
	6			5			4			3			2			1		
	OC	K	P	OC	K	P	OC	K	P	OC	K	P	OC	K	P	OC	K	P
1	0.27	15	10	0.41	20	10	0.39	90	10	0.47	890	30	0.24	1,380	20	0.35	1,170	60
2	0.02	10	10	0.02	10	10	0.05	60	10	0.15	1,160	30	0.06	1,460	40	0.09	1,200	45
3	0.02	10	10	0.02	10	10	0.04	190	10	0.04	1,170	20	0.05	1,360	40			
4	0.02	10	10	0.02	10	10	0.19	990	40									
5	0.02	10	10	0.02	10	10	0.21	1,350	55									
6	0.02	10	10	0.02	10	10	0.15	1,280	35									
7	0.02	10	10	0.02	10	10	0.11	1,250	35									
8	0.02	10	10	0.02	10	10	0.08	1,250	45									
9	0.02	10	10	0.02	10	10												
10	0.02	10	10	0.02	10	10												
11	0.02	10	10	0.02	10	10												
12	0.02	10	10	0.06	60	15												
13	0.02	10	10	0.20	300	30												
14	0.03	10	10	0.30	290	20												
15	0.04	20	20															

OC, organic carbon (%) by dry combustion in a Leco furnace. K and P, potassium (p.p.m.) and phosphorus (p.p.m.) by X-ray fluorescence. Limit of detection is ~10 p.p.m.

severely eroded and depth to B horizon varies accordingly, for example, in system 6 it ranges from 12 to >20 m.

Differences in soil morphology between the dune systems are reflected in the distribution of carbon, potassium and phosphorus down the profile (Table 1). The organic carbon profile indicates the development of A and B horizons. The potassium and phosphorus profiles reflect the weathering of feldspar and the removal of sesquioxide coatings from the quartz grains in the upper part of the profile. The unweathered sands contain virtually no apatite and phosphorus is held on the sesquioxide coatings.

Clay formation is limited by the very low content of weatherable minerals in the parent sands; the clay-size fraction in B horizons examined in systems 1–3 is <3%, in system 4 <7%, in system 5 <10% and up to 15% in system 6. Chlorite-vermiculite and kaolin dominate the clay fraction in the younger soils and kaolin and quartz in the older. The silt-size fraction is negligible (<1%) in the young systems 1–4, but may reach 7% in the top of the B horizon in systems 5 and 6, with negligible amounts above and below this position.

Stratigraphical and geomorphological relationships, and the differences in degrees of erosion, weathering and soil development show that the age sequence of the dunes at Cooloola extends over a long period. However, the absolute ages of these dune systems are unknown and no datable material has been recovered in the present study. A previous radiocarbon dating¹² of two pieces of wood embedded in the Cooloola sandmass near sea level indicated ages >40,000 and >45,000 yr BP from materials probably contaminated by modern carbon in groundwaters which issue at both sites. Additional dating of charcoal embedded in dune sands and a soil organic B horizon in overlying sands, both about 75 m above sea level, yielded ages of 39,000 ± 3,000 and 30,300 ± 800 yr BP¹². Although the latter site cannot be located in relation to the dune systems now recognized it does show that the underlying sands are of Pleistocene age.

Dune systems 1–3 transgress, and in places fill, parts of deeply eroded valleys in the older dune systems. The preservation of dune outlines without erosion breaks and the limited depth of weathering in systems 1–3 (Table 1) also indicate young deposits. The diminishing size of these systems and of the individual dunes towards the coast suggests successive deposition from a diminishing sand supply during a high stillstand of the sea⁷ as has occurred along the Australian coast during the Holocene¹⁵. Also, the degree of soil development on dunes in system 3 is similar to that observed in the dune at Triangle Cliff, Fraser Island, where a recent radiocarbon dating of charcoal at the base

of the dune shows that it is <3,880 yr BP¹⁶. Therefore, systems 1, 2 and 3 are most likely of Holocene age.

In dune system 4, erosion has breached the trailing arms and reduced most of the dune floors and, together with the increased depth of soil weathering (Fig. 2; Table 1), indicates a sizeable time break between it and system 3. Even longer intervals are indicated by the increases in erosion and weathering between systems 4 and 5 and between 5 and 6. The increasing degree of modification indicates that these systems are older than Holocene and the early radiocarbon dates¹² show that their ages extend beyond 40,000 yr BP.

In New South Wales, two distinct barrier systems of low sand ridges have been recognized⁹—an Outer Barrier dated as a Holocene feature deposited over the past 6,000 yr (ref. 15) and an Inner Barrier which has been related, by uranium series dating of corals it overlies, to high sea levels of the last Pleistocene interglacial, ~120,000–135,000 yr BP¹⁷. While Inner and Outer Barrier equivalence has not been clearly established for the high dune systems along the Queensland coast, the younger

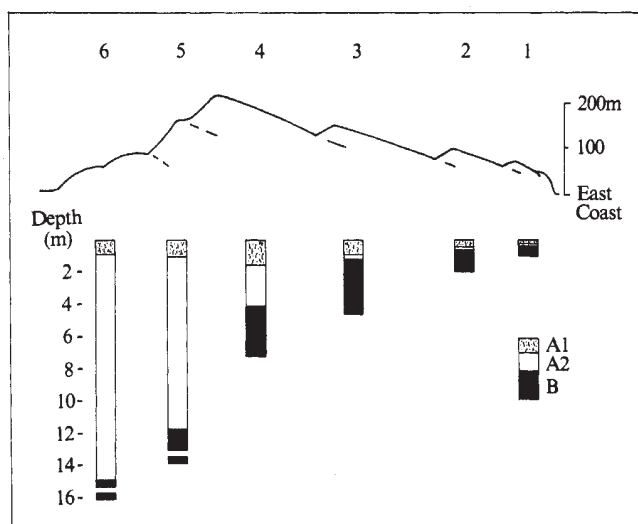


Fig. 2 Schematic cross-section of overlapping dune systems at Cooloola and podzol profiles associated with them. Numerals indicate dune systems, 1 being the youngest and 6 the oldest. The profiles represent soils at sites of least erosion for each system (dune floor in systems 1–4 and ridge crests in systems 5, 6). Thickness of B horizons in systems 5 and 6 is unknown.

systems (1–3) are probably equivalent to the Holocene transgressive parabolic dunes developed on parts of the Outer Barrier⁹, judging from the limited amount of erosion and weathering evident at both localities.

Equivalence between the older dune systems at Cooloola and the Inner Barrier Deposits is less certain because of the greater areas of old deposits at Cooloola and the evidence of more events than can be reasonably represented by a single unit. For example, while no coastal sand deposition older than the Inner Barrier has been recognized in New South Wales¹⁷, the coloured Teewah Sands are prominent in southern Queensland¹² and probably predate the Inner Barrier deposits. At Cooloola, the Teewah Sands underlie or form part of dune system 5 and at its northern end are underlain by older dune sands (system 6) which overlie a fossil beach. If system 6 deposits here equate with the Inner Barrier as suggested¹⁸, there seem to be more depositional events at Cooloola between the last interglacial and Holocene than is suggested by records elsewhere. Alternatively, system 5 at Cooloola can be regarded as the most likely equivalent of the Inner Barrier and system 6 as an older deposit as has been suggested for the fossil beach¹⁹.

In the Cooloola area, there is no firm indication of a change to markedly drier conditions during the last glaciation. The parabolic shape of the dunes in systems 1–4 and the lack of dune shapes characteristic of extensive bare sands suggest that this area has carried ground cover continuously since before the beginning of deposition of dune system 4. However, the evidence of extensive fluvial erosion in systems 4–6 suggests that they have experienced wetter conditions at some time in the past than those that now prevail or are believed to have occurred since the onset of the last glaciation²⁰.

The soil landscapes here and those on Fraser, North Stradbroke and Moreton Islands have been correlated¹⁰ and more recent field examinations indicate that sequences of podzols similar to that at Cooloola also occurs on these islands. Another such sequence is evident in the high dune systems further north at Shoalwater Bay (22.5° S) where at least four of the systems can be equated with those at Cooloola. To the south, soils on the parabolic dunes of the Outer Barrier⁹ in the Myall Lakes area (33° S) are also similar to those of at least two of the systems at Cooloola. The soils of the parabolic dune systems further to the north and south have not yet been correlated; however, air-photo patterns of the dunes at Cape Flattery (15° S) and Shelburne Bay (12° S) show similar dune forms and descriptions from King Island²¹ (40° S) suggest that one or more of the dune systems recognized at Cooloola are represented in these areas.

It is evident that the podzol sequence at Cooloola has developed over a considerable period including at least the last interglacial, last glaciation and postglacial periods, and podzol sequences on similar dune systems to the north and south can be correlated with it.

These podzol chronosequences provide unique opportunities for scientific study, both of changes that have occurred with time at one locality and of parallel changes that have occurred in temperate, subtropical and tropical areas. The sanddune systems at Cooloola are being intensively studied by the CSIRO Division of Soils, Brisbane, with collaboration from other institutions. Recognition of the dune systems and the podzol chronosequence has provided the framework for studies of erosion, nutrient movement, atmospheric accretion, dune hydrology, and ecology of vegetation, fungi and soil fauna—these investigations need extending to the north and south.

Received 22 September 1980; accepted 2 March 1981.

- Burgess, A. & Drover, D. P. *Aust. J. Bot.* **1**, 83–94 (1953).
- Dickson, B. A. & Crocker, R. L. *J. Soil. Sci.* **4**, 123–154 (1953).
- Franzmeier, D. P. & Whiteside, E. P. *Mich. Agr. Exp. Stat. Q. Bull.* **46**, 1–57 (1963).
- Cowie, J. D. N. *J. Geol. Geophys.* **6**, 268–280 (1963).
- Jenny, H., Arkley, R. J. & Schultz, A. M. *Madrono* **20**, 60–74 (1969).
- Bird, E. C. F. in *Barrier Islands* (ed. Schwartz, M. L.) 410–426 (1973).
- Walker, H. J. *Soil. Sci.* **119**, 3–19 (1975).
- Shakleton, M. J. & Opdyke, N. D. *Quat. Res.* **3**, 39 (1973).
- Thom, B. G. *Proc. R. Soc. N.S.W.* **98**, 23–36 (1965).
- Thompson, C. H. *Proc. R. Soc. Qd* **86**, 109–120 (1975).

- Thompson, C. H. & Hubble, G. D. *Proc. int. Conf. CLAMATROPS*, Kuala Lumpur, 179–188 (1979).
- Coaldrake, J. E. *Proc. R. Soc. Qd* **72**, 101–116 (1962).
- Little, I. P., Armitage, T. M. & Gilkes, R. J. *Geoderma* **20**, 225–237 (1978).
- Brewer, R. & Thompson, C. H. *Proc. int. Conf. CLAMATROPS*, Kuala Lumpur (1979).
- Thom, B. G. & Chappell, J. *Search* **6**, 90–93 (1975).
- Grimes, K. G. *Qd Govt Min. J.* **80**, 79–82 (1979).
- Marshall, J. F. & Thom, B. J. *Nature* **263**, 120–121 (1976).
- Langford-Smith, T. *Search* **3**, 102–104 (1972).
- Ward, W. T., Little, I. P. & Thompson, C. H. *Paleogeogr. Paleoclimatol. Paleocol.* **26**, 305–316 (1979).
- Bowler, J. M., Hope, G. S., Jennings, J. N., Singh, G. & Walker, D. *Quat. Res.* **6**, 359–394 (1976).
- Jennings, J. N. *Geogr. J.* **123**, 59–70 (1957).

Climatically controlled variation of Sr and Mg in Quaternary planktonic foraminifera

Hans G. Cronblad & Björn A. Malmgren

Department of Geology, Stockholm University, Box 6801, S-113 86 Stockholm, Sweden

Despite studies on the distribution of Sr and Mg in shells of marine calcareous invertebrates, the question of possible seawater temperature control on the concentrations of these elements in carbonate material remains controversial. Some workers have shown that individuals of a species living in warmer waters are richer in Sr and Mg than those living in cooler waters^{1–7}. Others have found no such relationship^{8–12}. We report here in tests of two planktonic foraminiferal species that the Sr concentrations are strongly related to late Quaternary climatic oscillations. Mg concentrations also show correlations with palaeoclimate, but not consistently. Concentrations of these trace elements, especially of Sr, in planktonic foraminifera may, therefore, be valuable for analysing palaeoclimatic history.

We have examined two species, *Globigerina bulloides* and *Globorotalia inflata*, in two deep-sea sediment cores from the southern Indian Ocean: E49–19 from the central subAntarctic and E48–22 from the Subtropical Convergence (Table 1). *G. bulloides* is most abundant in subpolar waters, whereas *G. inflata* predominates in the transitional zone between subpolar and subtropical waters¹³. *G. bulloides* lives in 'intermediate waters' (upper 100 m of the water column) and *G. inflata* inhabits 'deep waters' (generally below 100 m)¹³.

Table 1 Data for Eltanin cores used in the study

Core	Lat. (S)	Long. (E)	Water depth (m)	Core length (cm)	Age of sequence (kyr)
E48–22	39°53.7'	85°24.6'	3,380	400	220
E49–19	43°53.2'	90°06.0'	3,036	550	300

Size-fractionation of samples may affect the oxygen isotope composition of planktonic foraminifera^{14,15}. Tests of small specimens tend to be isotopically lighter than tests of the larger size of the same species. A pilot study has shown that there are no major differences between Sr and Mg concentrations in the various size fractions in *G. bulloides* and *G. inflata* (Table 2). Therefore, we concentrated on the >250-μm fraction of these relatively large species. More than 150 specimens of each species were picked per sample.

Specimens were ultrasonically cleaned to remove sediment and clay contaminants, and analysed using standard atomic absorption spectrophotometric methods. Only those tests that were apparently free of extraneous material were included. The analytical precision (95% confidence) is ±14 p.p.m. for Sr and ±11 p.p.m. for Mg. These estimates were based on analyses of four replicate subsamples (two per species) from each of four samples. Sr and Mg variations were compared with palaeoclimatic indices established from 'total fauna' analyses and

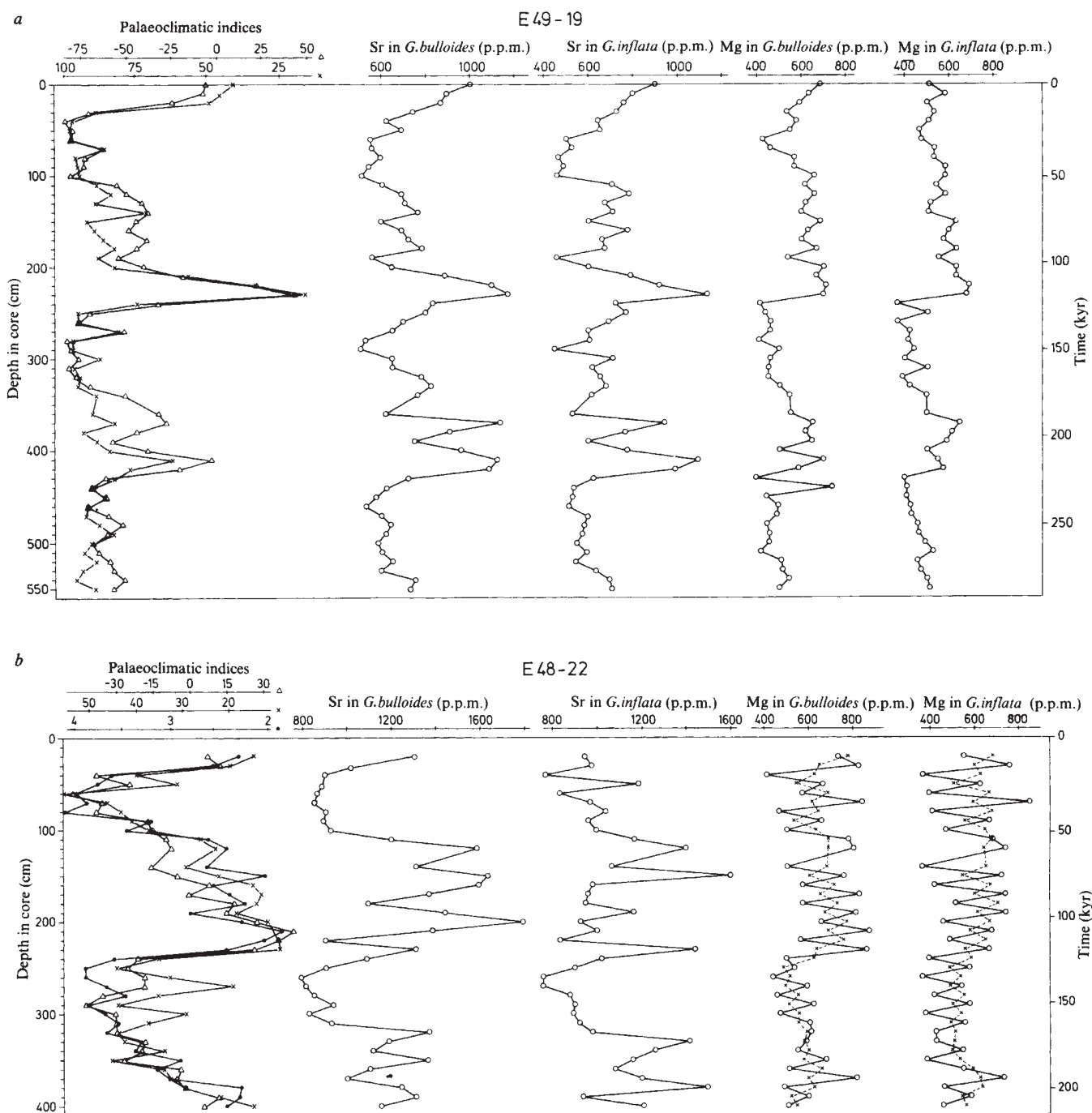


Fig. 1 *a*, Variation in Mg and Sr concentrations in *G. bulloides* in relation to two palaeoclimatic indices in core Eltanin 49-19 from the southern Indian Ocean. 'Warm' is to the right and 'cool' to the left. Chronology plotted on right is from Malmgren and Kennett¹⁷. *b*, Variation in Mg and Sr concentrations in *G. bulloides* and *G. inflata* in relation to three palaeoclimatic indices in core Eltanin 48-22 from the southern Indian Ocean. Dashed line curves for Mg show three-point moving averages. Chronology is from Williams¹⁶. Δ , 'Total fauna'; $\times$, % sinistral *N. pachyderma*; $\bullet$, $^{18}\text{O}/^{16}\text{O}$ ratios.

frequency in coiling proportions of *Neogloboquadrina pachyderma*, and also in E48-22 with a curve based on $^{18}\text{O}/^{16}\text{O}$ ratios in a planktonic foraminiferal species^{16,17}. A 'total fauna' index is defined as the difference between relative abundances of a transitional and a subAntarctic assemblage of planktonic foraminifera¹⁶.

In core E49-19, variations of the Sr concentrations in both species are remarkably similar in period and amplitude to the palaeoclimatic curves (Fig. 1). Mg concentrations are less variable and show less pronounced peaks, but the curves are clearly in phase with the palaeoclimatic oscillations. Cross-correlation

analyses indicate that all curves have a very significant correlation (Table 3). Both elements are more enriched during warm climates (interglacial episodes) than during cool climates (glacial episodes).

Core E48-22 also shows good agreements between the Sr patterns in both species and the palaeoclimate (Fig. 2). Cross-correlation analyses between Sr and palaeoclimatic changes indicate that correlations in *G. bulloides* are highly significant (Table 3). Sr in *G. inflata* is correlated with $^{18}\text{O}/^{16}\text{O}$ ratios, but not with the other curves. The Mg concentrations in *G. bulloides* are also correlated with palaeoclimatic indices, whereas those in

Table 2 Sr and Mg compositions in *G. bulloides* and *G. inflata* from different size fractions

Size fraction (µm)	52 cm				122 cm			
	<i>G. bulloides</i>		<i>G. inflata</i>		<i>G. bulloides</i>		<i>G. inflata</i>	
	Sr (p.p.m.)	Mg (p.p.m.)	Sr (p.p.m.)	Mg (p.p.m.)	Sr (p.p.m.)	Mg (p.p.m.)	Sr (p.p.m.)	Mg (p.p.m.)
175–250	900	630	1,150	650	1,450	810	1,070	760
250–351	880	630	1,180	620	1,490	810	1,090	740
>351	890	660	1,200	640	—	—	—	—

Size fraction (µm)	222 cm				362 cm			
	<i>G. bulloides</i>		<i>G. inflata</i>		<i>G. bulloides</i>		<i>G. inflata</i>	
	Sr (p.p.m.)	Mg (p.p.m.)	Sr (p.p.m.)	Mg (p.p.m.)	Sr (p.p.m.)	Mg (p.p.m.)	Sr (p.p.m.)	Mg (p.p.m.)
125–175	850	580	840	510	—	—	—	—
175–250	880	570	—	—	1,090	520	1,050	560
250–351	850	570	870	510	1,070	540	1,110	590
>351	890	560	850	500	1,090	530	1,070	550

Specimens were randomly picked from two cooler-water samples (52 cm and 362 cm) and two warmer-water samples (122 cm and 222 cm) from E48–22.

G. inflata are not. Small-scale fluctuations in this element are greater than in E49–19, but plots of three-point moving averages (Fig. 2) bring out these palaeoclimatic signals in *G. bulloides* more clearly. The peculiar 'sawtooth' pattern is probably only an analytical artefact.

Various hypotheses have been suggested for the calcification mechanism in marine calcareous organisms^{18–21}. The observed trends indicate that the biomineralization in planktonic foraminifera is, at least partly, controlled by seawater temperature. It has been shown empirically²² that substitution of exchangeable elements in the molecular lattice is greater at higher temperatures. Other environmental parameters may also play a part in the calcification process. The uptake of trace elements is also partly controlled by genetical differences among organisms. Differences in trace element concentrations among various planktonic foraminiferal species have been noted^{9,10}, and are apparent in our material (Fig. 1).

Sr and Mg concentrations in *G. bulloides* and *G. inflata* are lower on average in cores from cooler waters (E49–19), and still lower in samples from cool subAntarctic waters (compare tops

of cores in Fig. 1 and core-top data in Fig. 2). This suggests that temperature-related gradients exist in Sr and Mg concentrations. If they did, the down-core variations may reflect climatically induced periodic expansions and contractions of the cooler water masses to the south of the core sites. Despite the possible existence of similar latitudinal gradients in both elements, palaeoclimatic correlations are weaker in Mg. This may be because the Sr ions are more tightly bound to the calcite lattice²³, but could also be explained by the stronger influence of clay contamination on the Mg composition^{24,25}.

Within individual species of planktonic foraminifera, the degree of dissolution on the sea floor has been shown¹⁰ to affect the Mg concentration; a less dissolved sample is more enriched in Mg compared with a sample of the same species subjected to stronger dissolution. Differential susceptibility to dissolution between glacial and interglacial episodes could, therefore, also explain the trends in Mg. To determine dissolution effects on *G. bulloides* and *G. inflata*, we analysed recent subAntarctic core-top samples from the Maurice Ewing Bank (eastern Falkland Plateau), South Atlantic Ocean. No biogeographic

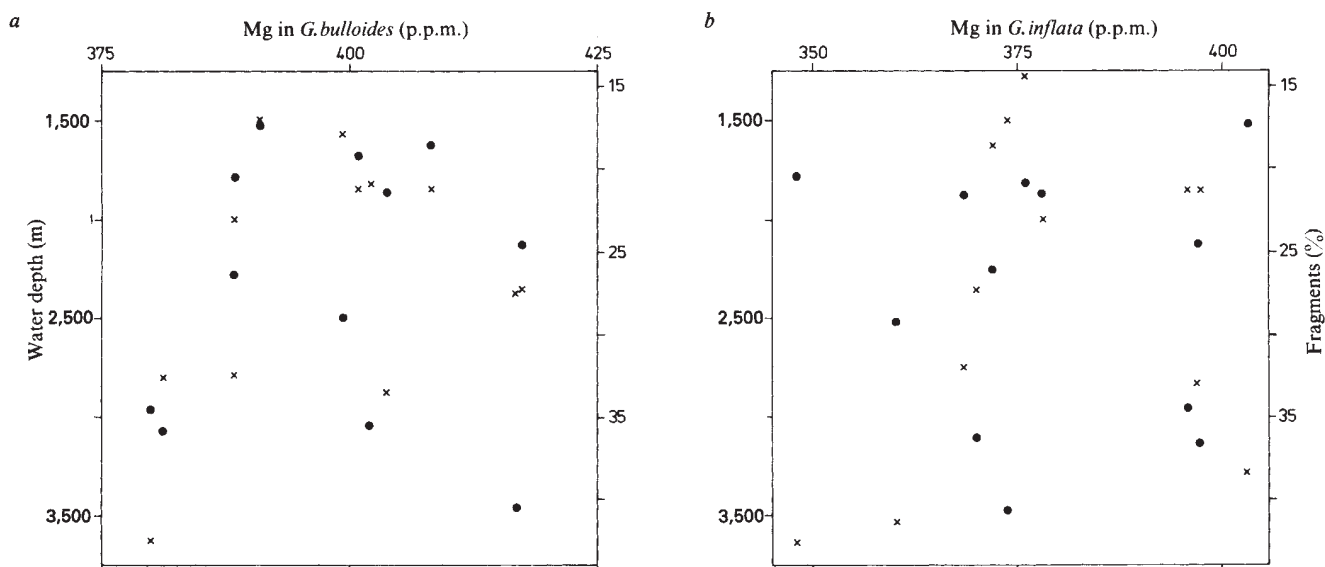


Fig. 2 Mg concentrations in *G. bulloides* (a) and *G. inflata* (b) in relation to water depth and degree of fragmentation of planktonic foraminiferal tests in cool subAntarctic core-top samples (Eltanin and Islas Orcadas) from the Maurice Ewing Bank, South Atlantic Ocean. Mg is not related to these parameters, which indicates that it is not affected by calcium carbonate dissolution as suggested previously¹⁰. ●, water depth; ×, % fragments.

Table 3 Cross-correlation coefficients among trace element compositions in *G. bulloides* and *G. inflata*, and palaeoclimatic indices in cores E49-19 and E48-22

E49-19		Total fauna	% Sinistral <i>N. pachyderma</i>	
Sr in <i>G. bulloides</i>	0.80*		-0.70*	
Sr in <i>G. inflata</i>	0.77*		-0.69*	
Mg in <i>G. bulloides</i>	0.64*		-0.51*	
Mg in <i>G. inflata</i>	0.57*		-0.43*	
E48-22		Total fauna	% Sinistral <i>N. pachyderma</i>	¹⁸ O/ ¹⁶ O
Sr in <i>G. bulloides</i>	0.54*		-0.50*	-0.69*
Sr in <i>G. inflata</i>	0.19		-0.13	-0.38†
Mg in <i>G. bulloides</i>	0.43†		-0.36†	-0.44‡
Mg in <i>G. inflata</i>	0.30		-0.27	-0.29

These coefficients were determined for lags of zero, where correlations are normally at a maximum.

* $P \leq 0.001$; † $P \leq 0.05$; ‡ $P \leq 0.01$.

differences exist over this narrow latitudinal belt (50–51°S), and differences in Mg should be due to dissolution. The Mg content does not change with water depth, and is also not related to the degree of fragmentation of the planktonic foraminiferal tests (Fig. 3). The degree of fragmentation has been shown^{26,27} to represent a dissolution index in marine sediments. Therefore, it is less likely that the Mg signals in the cores are affected by dissolution. Likewise, no dissolution effects were found on Sr, which is in agreement with previous observations²⁸.

The results indicate that Sr concentrations in *G. bulloides* and *G. inflata* closely follow the late Quaternary climatic variations. The material available allowed only single analyses per sample; the precision of individual sample estimates could be improved by replicate analyses. Also, with further refinements of laboratory techniques, much of the 'noise' in the Sr curves might be reduced, which could aid in clarifying the palaeoclimatic signals. If so, Sr concentrations in these and other planktonic foraminiferal species, are a new independent means for palaeoclimatic analysis of Quaternary and, possibly, older sediments.

This research was supported by grant G4076-101 from the Swedish Natural Science Research Council. We thank Drs M. L. Bender, J. P. Kennett, F. T. Manheim and L. Sjöberg for helpful comments, and D. S. Cassidy for providing the material from the South Atlantic. Chemical analyses were made by Adnan Fahmy, Stockholm University.

Received 6 October 1980; accepted 23 February 1981.

- Clarke, F. W. & Wheeler, W. C. *U.S. geol. Surv. Prof. Pap.* **124**, 1–62 (1922).
- Chave, K. E. *J. Geol.* **62**, 266–283 (1954).
- Chave, K. E. *J. Geol.* **62**, 587–599 (1954).
- Blackmon, P. D. & Todd, R. J. *Paleont.* **33**, 1–15 (1959).
- Dodd, J. R. *Geochim. cosmochim. Acta* **29**, 385–398 (1965).
- Schopf, T. J. M. & Manheim, F. T. *J. Paleont.* **41**, 1197–1225 (1967).
- Zolotarev, V. N. *Geochim. Int.* **11**, 347–352 (1975).
- Krinsley, D. *Micropaleontology* **6**, 297–300 (1960).
- Savin, S. M. & Douglas, R. G. *Bull. Geol. Soc. Am.* **84**, 2327–2342 (1973).
- Bender, M. L., Lorens, R. B. & Williams, D. F. *Micropaleontology* **21**, 448–459 (1975).
- Ragland, P. C., Pilkey, O. H. & Blackwelder, B. W. *Chem. Geol.* **25**, 123–134 (1979).
- Windom, H. L. & Smith, R. G. *Southeastern Geol.* **13**, 275–281 (1971).
- Bé, A. W. H. in *Oceanic Micropaleontology* Vol. 1 (ed. Ramsay, A. T. S.) 1–100 (Academic, London, 1977).
- Emiliani, C. *Science* **173**, 1122–1124 (1971).
- Berger, W. H., Killingley, J. S. & Vincent, E. *Oceanol. Acta* **1**, 203–216 (1978).
- Williams, D. F. *Mar. Micropaleont.* **1**, 363–375 (1976).
- Malmgren, B. A. & Kennett, J. P. *J. Paleont.* **52**, 1195–1207 (1978).
- Towe, K. M. & Lowenstam, H. A. *J. ultrastruct. Res.* **17**, 1–13 (1967).
- Travis, D. F. *Ann. N.Y. Acad. Sci.* **109**, 177–245 (1963).
- Harrison, P. M. in *Iron Metabolism—An International Symposium*, 40–56 (Springer, Berlin, 1964).
- Reimann, B. E. *F. Cell. Res.* **34**, 605–608 (1965).
- Krauskopf, K. *Introduction to Geochemistry* (McGraw-Hill, New York, 1967).
- Curtis, C. D. & Krinsley, D. *Geochim. cosmochim. Acta* **29**, 71–84 (1965).
- Kilbourne, R. T. & Sen Gupta, B. K. *Geol. Soc. Am., Southeastern Sect., 22nd A. Meet. (Abstr.)* **5**, 408–409 (1973).
- Jørgensen, N. O. *Bull. geol. Soc. Denmark* **24**, 299–325 (1975).
- Thiede, J. *Bull. geol. Soc. Am.* **84**, 2749–2754 (1973).
- Thunell, R. C. *Quat. Res.* **6**, 281–297 (1976).
- Lorens, R. B., Williams, D. F. & Bender, M. L. *J. sedim. Petrol.* **47**, 1602–1609 (1977).

Intra-sex and inter-sex sibling interactions as sex ratio determinants

P. D. Taylor

Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada K7L 3N6

Fisher¹ produced the first general argument that a random-mating sexually-reproducing population should devote equal reproductive resources to the production of male and female offspring, while Hamilton² considered several unusual situations in which the expenditure of reproductive resources should be biased towards one sex or the other. His 'host' model has been widely cited^{3–6} and at least three factors proposed as the cause of the sex ratio bias: local mate competition, sib-mating and inbreeding. The problem of which of these factors is the real cause of the biased sex ratio seems important in view of the large number of apparently different but often closely related sex ratio models which are now being proposed. It would be useful to have ways of classifying them according to the presence of different factors which affect the equilibrium sex ratio. Here, I present a general equation (equation (2)) for differential change in fitness which exhibits the equilibrium sex ratio as achieving a balance between the factor contained in Fisher's standard grandchild argument¹ and a second factor resulting from intra-sex sibling interactions over reproductive resources. In particular, I analyse Hamilton's host model² for the presence of these different factors.

I will present the model for the special case of discrete non-overlapping generations. Although many of the phenomena I wish to study lie outside this special case, the formulation of the model is complicated by the mixing of generations, and it seems best to start with a simple situation. We divide the life cycle of the organism into two phases. Phase 1 encompasses the time from birth to the point at which some significant sex-specific behaviour begins, and phase 2 continues from there to the birth of the next generation. The purpose of the model is to relate the effects of sex-specific sib interactions in phase 2 to the equilibrium sex ratio.

Suppose in the population a typical mated female produces a total of x^* daughters and y^* sons at the end of phase 1. Consider a rare female who manages to produce x daughters and y sons. Her fitness in the $*$ -population will be

$$w(x, y) = D(x, y)x + S(x, y)y \quad (1)$$

where D is her expected ultimate genetic contribution through daughters per daughter and S is her expected contribution through sons per son. If at the end of phase 1, the offspring disperse at random into the general population, with no further non random sib interaction, then $D(x, y)$ and $S(x, y)$ are constants equal to their values D^* and S^* at (x^*, y^*) ; but otherwise they may not be constant.

Constraints operating during phase 1 will restrict the possible values of x and y in the population to a one-parameter family parameterized by the proportion r of reproductive resources allocated by the mother to the production of sons. In the simple case of a linear interchange between sons and daughters, with both sexes equally expensive, we have $x = K(1-r)$ and $y = Kr$, where K is the litter size. With this parameterization, w , D and S become functions of r and the value of dw/dr at $r = r^*$ (the value giving x^* and y^*) is a measure of the evolutionary pressure for an increase in r^* . If $dw/dr > 0$ we expect the population to evolve towards an increase in the proportion of resources allocated to the production of sons; if $dw/dr < 0$, this proportion should decrease. Now, differentiating equation (1),

$$\frac{dw}{dr} = \left[D^* \frac{dx}{dr} + S^* \frac{dy}{dr} \right] + \left[\frac{\partial D}{\partial x} \frac{dx}{dr} + \frac{\partial D}{\partial y} \frac{dy}{dr} + \frac{\partial S}{\partial x} \frac{dx}{dr} + \frac{\partial S}{\partial y} \frac{dy}{dr} \right] \quad (2)$$

where everything is evaluated at r^* . The second term on the right measures the effect of sibling interactions in phase 2 on the mother's reproductive success, and it vanishes in the absence of these. Let us interpret the four summands of this term.

The partial derivative $\partial D/\partial x$ measures the change in the quantity 'fitness through daughters per daughter' per extra daughter, with number of sons, y , held constant. Thus, it measures the influence of extra daughters on the reproductive success of each daughter. For example, competition among daughters for reproductive resources during phase 2 would tend to make $\partial D/\partial x < 0$. Similarly, the analogous term $\partial S/\partial y$ measures absence of competition among brothers for reproductive resources. The cross term $\partial D/\partial y$ measures the influence of extra sons on the reproductive success of each daughter and similarly $\partial S/\partial x$ measures the influence of extra daughters on the reproductive success of each son.

Let us consider some examples. Clark⁷ has studied African galagos in which young males disperse, but females remain within their home ranges and compete with their mother and sisters for reproductive resources (space, food). The resulting competition between sisters may contribute negatively to $\partial D/\partial x$. In MacDonald's foxes⁸, females form cooperative breeding groups in which non-breeding vixens seem to confer benefits on the mother. There is strong evidence that members of a group are closely related, contributing positively to $\partial D/\partial x$.

Where males enhance the reproductive success of their sisters, we may get a positive contribution to $\partial D/\partial y$. Woolfenden⁹ has shown that male helpers at the Florida scrub jay nest can have a significant effect on reproductive output. In all three of the above examples much of the interaction is between generations and may have important effects on the equilibrium sex ratio. However, the simple assumption of our model of non-overlapping generations restricts our attention to interaction within the same generation.

If competition between males for mates is more intense between brothers than between unrelated males, this contributes negatively to $\partial S/\partial y$. Partial sib-mating often has this effect, because brothers are competing for their sisters. Positive contributions to $\partial S/\partial y$ may result from cooperative behaviour among brothers. This is apparently the case, for example, for the lions studied by Bertram¹⁰, who often hunt together and protect one another.

Partial sib mating also contributes positively to $\partial S/\partial x$, since extra daughters will increase the reproductive success of the sons. It is not so easy to find other examples in which daughters affect the reproductive success of their brothers. Pickering¹¹ describes a parasitoid wasp in which larval sibs growing on the host pupa compete for food. There is evidence that female larvae behave less altruistically towards their brothers than their sisters. An extra sister will affect both male and female reproductive success, but the effect on males may be greater.

To make specific calculations, we must find expressions for D and S in terms of x , y , x^* and y^* . In the above example of Pickering, the organisms are haplodiploid, and the resulting genetic asymmetry between males and females makes it difficult to find comparable measures for D and S . In this case special arguments are required^{6,12}. The calculation of equilibrium sex ratio is further complicated by a conflict between mother and daughter over sex ratio^{3,13}. In our calculations I will assume a diploid population. In this case, D and S can be calculated using the common measure of mated females; each mated daughter counts 1 and each female mated exclusively by a son counts 1 (with shared matings discounted by the appropriate factor). From this method of counting it follows that $D^*x^* = S^*y^*$.

The equilibrium value of r^* is obtained by setting dw/dr , given by equation (2), equal to zero. An examination of the various terms of equation (2) at equilibrium will display the different forces (on the value of r^*) which are held in balance at equilibrium.

As a normative example, consider Fisher's standard situation¹ in which offspring disperse after phase 1 and mate at random. Then $D(x, y) = D^*$, $S(x, y) = S^* = D^*x^*/y^*$, $x = K(1-r)$ and $y = Kr$. Therefore, all partial derivatives of D and S vanish, and equation (2) becomes

$$-D^*K + D^*K(1-r^*)/r^*$$

which, when set equal to zero, gives the equilibrium value $r^* = \frac{1}{2}$. We will refer to the first bracket in equation (2) as the Fisher term, as this is the term which must vanish at equilibrium in his standard model.

In Hamilton's model², mated females collect on hosts in random groups of size N to produce and raise K offspring each. At the start of phase 2, the KN offspring mate at random among themselves, and the mated females go out into the general population to form new host groups and start the cycle again. Because a mutant female is rare we can assume the other females in her host group are normal. Hence, $D(x, y) = D^*$ and $S(x, y) = D^*(x + (N-1)x^*)/(y + (N-1)y^*)$. With these expressions both partial derivatives of D vanish, $\partial S/\partial x$ is positive because of the sib-mating and $\partial S/\partial y$ is negative because of male reproductive competition resulting from the sib-mating. Then equation (2) has the form

$$D^* \left[\frac{dx}{dr} + \frac{x^*}{y^*} \frac{dy}{dr} \right] + \left[\frac{\partial S}{\partial x} \frac{dx}{dr} + \frac{\partial S}{\partial y} \frac{dy}{dr} \right]$$

Since $dx/dr < 0$ and $dy/dr > 0$ both terms in the second bracket are negative, and both act in the same way on r^* . At equilibrium the Fisher term must be positive, requiring a female-biased sex-ratio. With $x = K(1-r)$ and $y = Kr$, the equilibrium ratio turns out to be $r^* = (N-1)/2N$. Thus, the female bias, which is often loosely thought to result from sib-mating, actually results from two different forces: the effect of extra daughters on male mating success, and the effect of fewer sons on male reproductive competition.

Finally, the general model applies equally well to the model of Charnov¹⁴ for barnacles and Fischer¹⁵ for coral reef fish, of sex allocation in hermaphrodites. (Replace mated female by individual, daughter and son by egg and sperm, and the process of mating by that of fertilization.) Each individual delivers its sperm to fertilize the eggs of N other individuals, and has its own eggs fertilized by, on average, N other individuals. As before a linear tradeoff is assumed between production of eggs x and sperm y . Thus $D(x, y) = D^*$ and $S(x, y) = D^*Nx^*/(y + (N-1)y^*)$. This is formally the same as Hamiltonian's model, except that we have removed the possibility of sib-mating, and $\partial S/\partial x$ is now zero. However, $\partial S/\partial y$ is the same as before, again forcing the Fisher term to be positive, creating a female-biased equilibrium ratio, this time, only through the effect of fewer sons on male reproductive competition.

The significance of equation (2) is that it resolves the evolutionary pressure on r^* into a sum of two factors which are held in balance at equilibrium: the standard Fisher factor and a second factor containing the effect of intra- and inter-sex sibling interactions. This second factor again resolves into four terms. Different sex ratio models can be compared according to the presence of these factors. Sib-mating (or selfing in hermaphrodites) or more generally mating between relatives, often contributes to these factors, but the inbreeding which results, in itself, has no role in determining sex ratio.

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. I thank Michael Bulmer for many discussions which sharpened my understanding of the causes of sex ratio bias and George Williams for valuable comments on the first draft.

Received 18 August 1980; accepted 26 February 1981.

1. Fisher, R. A. *The Genetical Theory of Natural Selection* (Oxford University Press, 1929).
2. Hamilton, W. D. *Science* **156**, 477–488 (1967).
3. Trivers, R. L. & Hare, H. *Science* **191**, 249–263 (1976).
4. Stenseth, N. C. *Oikos* **30**, 83–89 (1978).
5. Maynard Smith, J. *The Evolution of Sex* (Cambridge University Press, 1978).
6. Taylor, P. D. & Bulmer, M. G. *J. theor. Biol.* **86**, 409–419 (1980).
7. Clark, A. B. *Science* **201**, 163–165 (1978).
8. MacDonald, D. W. *Nature* **282**, 69–71 (1979).
9. Emlen, S. T. in *Behavioural Ecology* (eds Krebs, J. R. & Davies, N. B.) 245–281 (Blackwell, Oxford, 1978).
10. Bertram, B. C. R. in *Behavioural Ecology* (eds Krebs, J. R. & Davies, N. B.) 64–96 (Blackwell, Oxford 1978).
11. Pickering, J. *Nature* **283**, 291–292 (1980).
12. Benford, F. A. *J. theor. Biol.* **72**, 701–727 (1978).
13. Bulmer, M. G. & Taylor, P. D. *Heredity* (in the press).
14. Charnov, E. L. *Mar. Biol. Lett.* **1**, 269–272 (1980).
15. Fischer, E. A. *Am. Nat.* **117**, 64–82 (1981).

Retinal projection in a non-visual area after bilateral tectal ablation in goldfish

S. C. Sharma

Department of Ophthalmology, New York Medical College, Valhalla, New York 10595, USA

If, in the adult goldfish, one optic tectum is ablated, the regenerating optic axons from the contralateral retina innervate the remaining tectum, where they form a retinotopically ordered map^{1–3}. The pathway for this induced ipsilateral projection¹ coincides with many of the pathways which normally connect the two tecta, but early in regeneration the optic fibres also enter non-visual centres to which there are degenerating tectal efferent pathways to follow⁴. We have therefore now investigated the fate of regenerating optic axons in goldfish from which both optic tecta have been removed; they are found to innervate non-visual centres, where again they generate a retinotopic map.

Fifteen adult goldfish (10–12 cm long and 2.5 yr old at the time of surgery) were anaesthetized with 0.5% MS222. The tectum was ablated bilaterally and extreme care taken not to damage the diencephalon, which joins the tectum at its rostral pole. Animals were allowed to survive between 380 and 620 days after tectal removal before electrophysiological mappings of the retinal projections were made. In five animals, the right eye was injected with 10 μ l of [3,4-³H]proline (specific activity 50 Ci mmol⁻¹) 1 day before mapping; the brains of these animals were processed for autoradiography. In five other animals, after visuotopic mapping, the right optic nerve was cut and 50% horseradish peroxidase (HRP) solution (Sigma type VI) applied to the cut end. The retinal projection of the remaining five animals was studied anatomically, using HRP orthograde transportation after injection of 5–7 μ l of 30% HRP intracocularly. In the HRP studies, animals were killed 1–3 days later and processed for HRP reaction⁵.

Autoradiographic analysis showed that, in addition to a high concentration of termination sites in the primary diencephalic visual centres, the optic fibres terminated in most tectal efferent areas^{6,7}. In normal fish these areas do not receive direct retinal fibre terminations^{8,9} and include the tori semicircularis, nucleus isthmi, nucleus of the rostral mesencephalic tegmentum, dorsolateral tegmental nucleus and nucleus reticularis superior of the hindbrain. In addition, radioactive grains were seen in the rostral valvula cerebelli.

Orthograde transport of HRP studied in 10 animals basically confirmed the autoradiographic identification of target centres of regenerating optic axons (Fig. 1). Most optic fibre fascicles in the regenerated optic nerve appear linearly arranged. The optic fibre fascicles separate from the ventral part of the optic tract at the diencephalic level and travel caudally before arborizing. The first bundle of optic fibres detached itself from the nerve at its rostral ventral position then the next dorsal bundle did the same. The uppermost portion of the nerve travelled farthest and apparently terminated in the valvula cerebelli (Fig. 1b, c). This pattern of axonal trajectories from the optic nerve is similar to

that observed in normal goldfish (S.C.S., unpublished observation).

The visuotopic projection of either the right or left eye was determined in 10 animals by electrophysiologically mapping the area directly beneath the ablated tectum. This area includes the valvula cerebelli medially and the torus semicircularis laterally and ventrally. The visuotopic projection represented in Fig. 2 was essentially similar to that seen in eight other animals and was confined to the rostro-medial portion of the contralateral valvula cerebelli. The projection map in the remaining animal was incomplete. The representation of the superior contralateral visual field was similar in its extent to a normal visuotopic projection onto the contralateral tectum. The receptive field sizes of the multi-units were in the 10–15° range, which is similar to the size of normal retinotectal units.

In three animals, in addition to the projection to the valvula cerebelli, large receptive fields (60–80°) were recorded from the torus semicircularis and were mostly from the inferior visual field. It was not possible to record any meaningful visuotopic maps from the torus. Large receptive field units were also recorded from the deeper tegmental areas.

These results imply that, in the absence of both optic tecta, the optic axons can regenerate to non-visual centres and generate a retinotopic projection which approximates to the normal map. As the retinotopic map represents presynaptic optic fibre terminations, it is not certain whether fibres formed synaptic connections with cells in anomalous areas. Anomalous connections to non-visual targets following one tectal removal have been reported in various vertebrates^{10–13}.

The valvula cerebelli extend rostrally into the optic ventricle and lie apposed to the ventral surface of the dorsal tectum

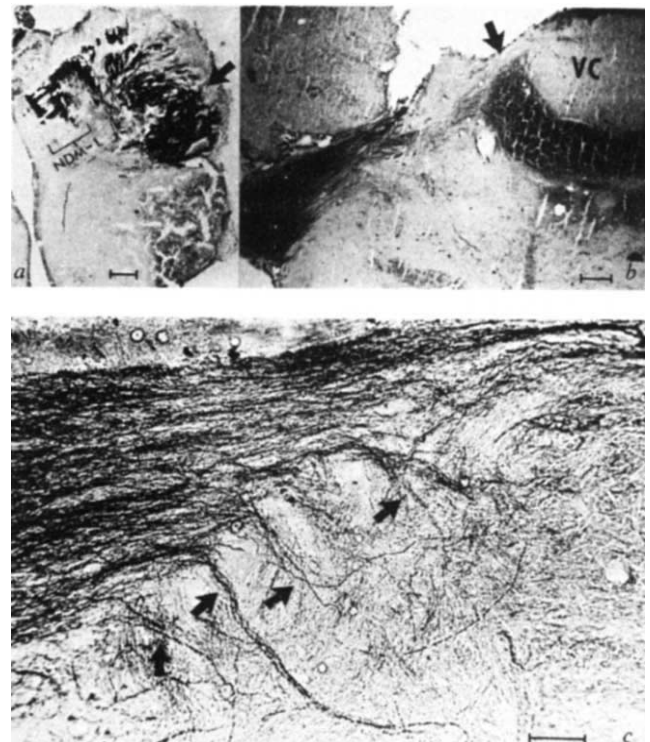


Fig. 1 *a*, Coronal section through the diencephalon showing contralateral retinal projection in a fish 380 days after tectal ablation where the eye was injected with HRP. The medial diencephalic retinal target areas (nucleus dorsalis medialis and lateralis, NDM, NDL are heavily labelled. The arrow points to a neuroma (a misleading entity as neuroma does not appear in sagittally sectioned brains). Scale bar, 100 μ m. *b*, Sagittal section through the brain of a fish 390 days after tectal ablation. The arrow points to the fibre entering the valvula cerebelli (VC). Scale bar, 100 μ m. *c*, High power photomicrograph of the optic nerve from *b*. The arrows point to optic fascicles separating from the nerve. Scale bar, 100 μ m.

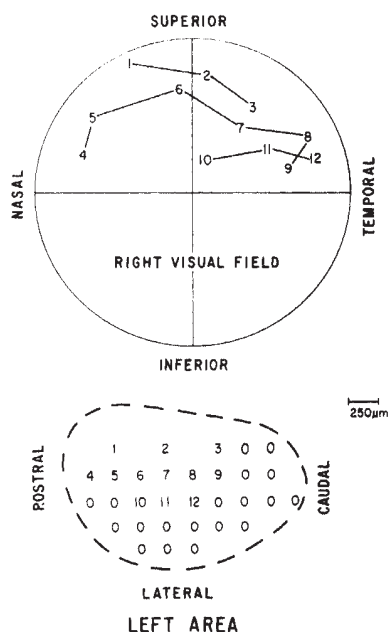


Fig. 2 The visual projection map from the right eye to the left portion of the valvula cerebelli, marked in the diagram as left area. Each number in the visual field shows the position of the stimulus that evoked maximum potential, recorded by an electrode at the position indicated by the same number on the left area. Empty circles represent electrode positions from which visually evoked potentials could not be recorded.

at the level of the diencephalic tegmental junction. Similarly, the rostral extent of the torus semicircularis begins at the level of the posterior commissure and is apposed to the ventral surface of the ventrolateral tectum. Removal of optic tecta is followed by injury and deafferentation to these structures, with the regenerating optic fibres entering them at their rostral poles. It is suggested here that innervation of the regenerating optic axons is facilitated by degenerating debris.

These results contradict the recent findings in frogs of Bohn and Stelzner¹⁴, who, after unilateral tectal ablation and lesion in the isthmal region, showed that regenerating optic axons did not innervate the non-visual areas deafferented by isthmal transection. Note that in Bohn and Stelzner's study, one optic tectum was always left intact¹⁴; when a tectal tissue is left intact, optic fibres preferentially innervate it¹⁵.

In normal goldfish, most optic axons terminate in the contralateral tectum. However, in the absence of both tecta, regenerating optic axons innervate non-visual areas and apparently form a topographically ordered map. Hence, it could be concluded that the tectum is not necessary for the formation of an ordered projection by the optic axons. The present results support the possibility that topography is intrinsic to the optic axons and that the formation of an ordered map may be guided by the presence of general axial cues (that is, rostral-caudal, medio-lateral) within the brain tissue.

This work was supported by NIH grant NEI01426. S.C.S. is the recipient of a Research Career Development Award from NIH-NEI.

Received 25 September 1980; accepted 2 February 1981.

1. Easter, S. S., Schmidt, J. T. & Leber, S. M. *J. Embryol. exp. Morph.* **45**, 145-159 (1978).
2. Levine, R. & Jacobson, M. *Expl. Neurol.* **43**, 527-538 (1974).
3. Sharma, S. C. *Expl. Neurol.* **41**, 661-669 (1973).
4. Lo, R. Y. S. & Levine, R. L. *J. comp. Neurol.* **191**, 295-314 (1980).
5. Mesulam, M. M. *J. Histochem. Cytochem.* **24**, 1273-1280 (1976).
6. Grover, B. & Sharma, S. C. *J. comp. Neurol.* **184**, 435-454 (1979).
7. Grover, B. & Sharma, S. C. *J. comp. Neurol.* **196**, (in the press).
8. Reperant, J., Lemire, M., Miceli, D. & Peyrichoux, J. *Brain Res.* **118**, 123-131 (1976).
9. Sharma, S. C. *Brain Res.* **39**, 213-223 (1972).
10. Bassinger, J., Lund, R. D. & Miller, B. *Expl. Neurol.* **54**, 369-382 (1977).
11. Kalil, R. E. & Schneider, G. E. *Brain Res.* **100**, 690-698 (1975).
12. Kicliter, E., Misentone, L. J. & Stelzner, D. J. *Brain Res.* **82**, 293-297 (1974).

13. Schneider, G. E. & Jhaveri, S. R. in *Plasticity and Recovery of Function in the Central Nervous System* (eds Stein, D. G., Rosen, J. J. & Butters, N.) 65-109 (Academic, New York, 1974).
14. Bohn, R. C. & Stelzner, D. J. *J. comp. Neurol.* **190**, 763-779 (1980).
15. Sharma, S. C. *Brain Res.* **183**, 453-457 (1980).

Birds can overcome the cardenolide defence of monarch butterflies in Mexico

Linda S. Fink* & Lincoln P. Brower†

* Webster Center for the Biological Sciences, Amherst College, Amherst, Massachusetts 01002, USA

† Department of Zoology, University of Florida, Gainesville, Florida 32611, USA

Flocks of black-backed orioles (*Icterus abeillei* Lesson) and black-headed grosbeaks (*Pheucticus melanocephalus* Swainson) eat several hundred thousand monarch butterflies (*Danaus plexippus* L.) in the dense overwintering colonies in central Mexico, and in 1979 were responsible for over 60% of the butterfly mortality at several sites¹. Such predation is unusual because, during larval development the aposematically coloured monarch butterfly sequesters cardenolides from its milkweed foodplants (Asclepiadaceae)^{2,3}. These bitter-tasting heart poisons cause vomiting in 12 species of birds in 9 families (refs 4-9, and L. P. Brower, unpublished observation), although the domestic chicken, Japanese quail, hedgehog, mouse and sheep^{8,10,11} have been shown to be insensitive to their emetic effects. Extensive predation of monarch butterflies by birds has never been observed except in Mexico. We report here that the Mexican butterflies are weakly emetic, and that taste discrimination by orioles and cardenolide insensitivity of grosbeaks allow these birds to feed freely on monarch butterflies.

To verify that a substantial proportion of the Mexican butterflies contain cardenolides, we collected 198 of these in January 1980 from Site Alpha¹ and stored them on dry ice before spectrophotometric assay^{12,13}. In common with wild populations from Massachusetts and California¹⁴, they varied widely in total cardenolide content (Fig. 1).

Our first hypothesis was that the birds taste each butterfly captured, release the emetic (bitter) ones, and kill and eat only the individuals containing little or no cardenolide¹⁵. Observing the birds through binoculars, we saw orioles attack 98 butterflies and release 75 without apparent harm; grosbeaks similarly released 65 of 92. Because orioles and grosbeaks never eat the butterfly wings we were able to measure the cardenolide content of butterflies killed and partly eaten by the birds. We collected damaged butterflies from the forest floor during feeding bouts from 7 to 19 January 1980, and spectroassayed one fore and one hind wing of each. The cardenolide concentrations of the wings of 177 butterflies damaged by grosbeaks, 142 damaged by orioles and 150 undamaged (control) butterflies are shown in Fig. 2. The frequency distributions of the three samples are very similar which provides strong evidence that the birds kill monarch butterflies randomly with respect to cardenolide content.

Our second hypothesis, based on our observations that the birds eat only parts and drop large portions of the butterflies, was that the birds ate less of the more emetic individuals. Table 1 relates the cardenolide content of the 142 butterflies damaged by orioles to the degree to which they were eaten by the birds. Clearly orioles eat fewer of the butterflies with more cardenolide content ($P < 0.005$). Damage to thoraces showed the same relationship: mean cardenolide concentration of wings of butterflies with heavily damaged, slightly damaged and

* Present address: Section of Ecology and Systematics, Cornell University, Ithaca, New York 14853, USA.

† To whom all correspondence should be addressed.

undamaged thoraces was 32, 44 and 55 μg per 0.1 g, respectively (one way analysis of variance, $F_{2,138} = 4.8$, $P < 0.01$). Thus, although individual butterflies were not initially protected from predation by containing more cardenolide, once killed, the amount of butterfly eaten by an oriole was inversely proportional to cardenolide content. The oriole further reduces its cardenolide intake by stripping out the thoracic muscle and abdominal contents without eating the cuticle in which these

Table 1 Cardenolide concentrations (μg per 0.1 g dry weight of the wings) of monarch butterflies eaten to various degrees by black-backed orioles at the Site Alpha Mexican overwintering colony, January 1980

Damage category	% Of abdominal contents eaten	Sample size	Cardenolide concentration (μg per 0.1 g dry weight)		
			$\bar{x}$	s.d.	Range
a	91–100	29	20	30	0–106
b	51–90	43	31	41	0–156
c	11–50	32	49	50	0–204
d	1–10	25	57	58	0–234
e*	0	13	65	72	0–208
Entire sample		142	41	50	0–234

The butterflies are grouped into five categories of decreasing abdominal damage. The data indicate a highly significant (one way ANOVA $F_{4,141} = 3.91$, $P < 0.005$) inverse relationship between the cardenolide content of the butterflies and the extent to which orioles eat them. Duncan's multiple range test indicates that categories a and b differ from categories c–e ($P < 0.05$).

* Abdomen unharmed, thorax partly eaten.

compounds are more highly concentrated¹⁶. Table 2 shows a similar analysis of the 177 butterflies damaged by grosbeaks which indicates no overall relationship between the extent of damage and cardenolide content ($P = 0.32$). In contrast to the orioles, grosbeaks consume cuticle as well as internal tissues. Thus, the orioles are selective once they have killed the butterflies, whereas the grosbeaks kill and continue to eat the butterflies randomly with respect to cardenolide content.

We next tested the sensitivity of the orioles and grosbeaks to the presumed emetic effect of the butterflies. For comparison with previous data, we first used our standard blue jay emetic dose 50 (ED_{50}) assay in which predetermined doses of dry powdered butterflies are force-fed in gelatin capsules to jays³. (The ED_{50} is that dose at which there is a 50% probability that a bird will vomit.) In the first test we combined 24 butterfly abdomens from Site Alpha (January 1980 collection), and fed four jays increasing doses up to 590 mg of powder containing 236 μg of cardenolide—no bird vomited. This high dose, equivalent in weight to more than six abdomens, reached the upper physical limit of the test. To make a more potent force-feeding powder we screened butterflies for cardenolide content and selected those with substantial amounts. Two wings from each of 453 Site Alpha butterflies were tested qualitatively with the spectroassay reagents, and only the abdomens of those showing a positive reaction were used ($n = 158$). This screening increased the concentration of cardenolides in the force-feeding powder from 40 to 105 μg per 0.1 g. We assumed, as determined in the Massachusetts and California populations^{14,17}, that we did not differentially select for particular cardenolides. For 25 jays the ED_{50} was 323 μg of cardenolide per 85 g (95% confidence limits = 247–425 μg).

We next tested the rosebreasted grosbeak (*Pheucticus ludovicianus* L.) which is closely related to the black-headed grosbeak¹⁸. Six birds were force-fed the powder previously given to the jays in doses from 132 to 331 μg of cardenolide (that is, from 0.41 to 1.0 blue jay emetic units). None of these birds vomited. Two additional grosbeaks were force-fed abdomens of butterflies reared on *Asclepias curassavica* L., previously shown¹⁹ to contain cardenolides 7.5 times as emetic as those of the Mexican monarch butterflies (Jay ED_{50} 43 μg per 85 g). One bird was fed an intact abdomen, estimated by spectroassay of a sibling to contain 270 μg of cardenolide. The second was fed a capsule of powdered abdomens containing 700 μg (16 blue jay emetic units) of cardenolide. Neither bird vomited, although both showed other signs of poisoning^{5,20}.

In January 1981 at Site Alpha, 10 black-headed grosbeaks and 11 black-backed orioles were force-fed powder made from monarch butterflies collected in Santa Cruz, California, the jay ED_{50} of which was previously estimated to be ~75 μg (ref. 14).

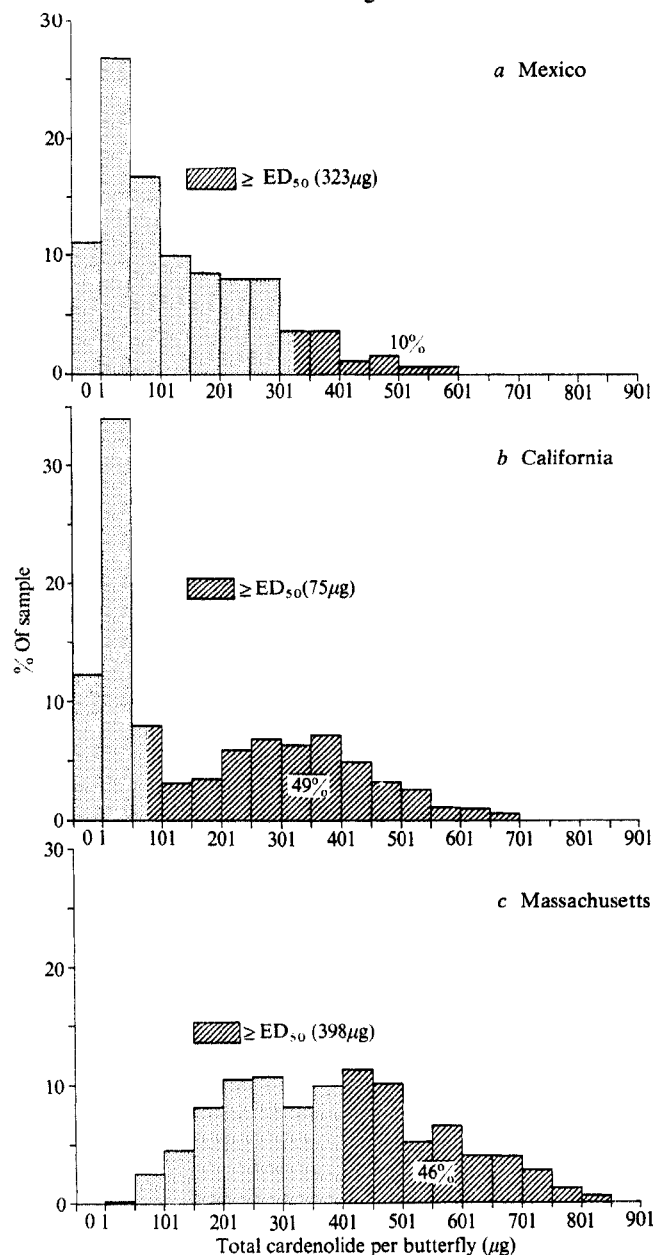


Fig. 1 a–c Compare total cardenolide per butterfly in overwintering monarch butterflies from Mexico (a) and California (b), and in autumn migrants from western Massachusetts (c). The hatched portions of the histograms are butterflies which contain sufficient cardenolide to cause a blue jay to vomit (that is, contain ≥ 1 blue jay ED_{50} unit per butterfly). Because the Mexico population combines weakly emetic cardenolides with a high proportion of low cardenolide butterflies, only 10% carry more than one emetic dose, in contrast to 49 and 46% of the California and Massachusetts butterflies. The potential protection of the Mexico population against sensitive bird predators is therefore much reduced. The data for the Massachusetts and California populations are calculated according to the method of Brower *et al.*¹³ from previously published absorbance values¹⁴. a, Site Alpha, 1980: $n = 198$, $\bar{x} = 128$, s.d. = 126, range 0–551. b, Muir Beach, 1971: $n = 351$, $\bar{x} = 165$, s.d. = 179, range 0–684. c, Hockanum, 1971: $n = 355$, $\bar{x} = 384$, s.d. = 172, range 34–850.

All birds were fed 130–140 mg of powder containing 146–157 μg of cardenolide, except one oriole that was fed half this dose. As controls, four additional orioles were force-fed 130–140 mg of ground Mexican butterfly abdomens, which contained low cardenolide content (35–38 μg). All 11 experimental orioles vomited, but none of the grosbeaks or control orioles became ill. These data show that the oriole is sensitive to the emetic effects of the cardenolide, but that the grosbeak is not.

The ED_{50} of the Mexican butterflies, combined with the data shown in Fig. 1a, shows that most of the Site Alpha monarch butterflies contain less than one emetic unit of cardenolide. The Mexico and California populations have a high proportion of butterflies with little or no cardenolide and contrast with the Massachusetts population in which no butterfly completely lacks poison. However, the cardenolides in both the Mexico and Massachusetts populations are much less emetic than those in the California butterflies. Consequently, the probability of any bird ingesting an emetic dose in Mexico is lower than in either of the other two populations (Fig 1b, c).

We offer three possible explanations for the high percentage of relatively non-emetic butterflies at Site Alpha. Total cardenolide decreases with time in overwintering monarch butterflies^{21,22}, and may result in increasing predation as the winter progresses. Second, Urquhart²³ has suggested that butterflies from different regions of the eastern breeding range may migrate to geographically distinct overwintering sites in Mexico. If this is correct, then other Mexican sites may contain more emetic butterflies reflecting regional differences in the local milkweed florals²⁴. However, heavy bird predation has been observed at all known sites¹. A third more general explanation results from preliminary TLC separations of the cardenolides in the Site Alpha butterflies. These suggest that a large proportion of the butterflies fed as larvae on *Asclepias syriaca* L. and/or *Asclepias speciosa* Torr., both of which contain low-potency cardenolides^{9,25}. Agricultural and land use changes have increased the abundances of both plants in North America^{24,26}.

Our finding that individual monarch butterflies containing high concentrations of cardenolides are not protected against two major bird predators raises questions about the part played by these compounds in the butterflies' defence. It is possible that

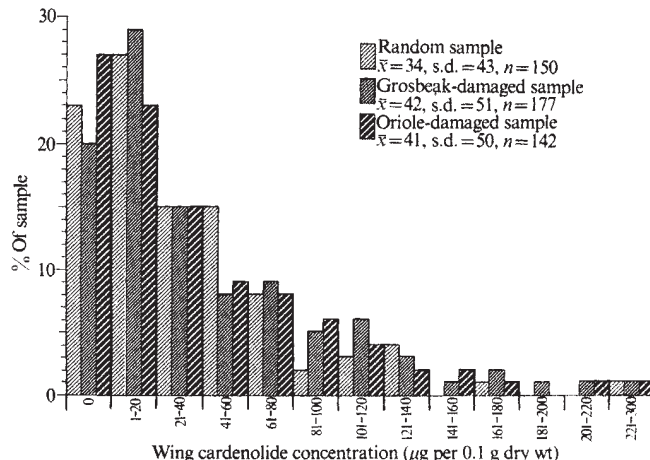


Fig. 2 Frequency distributions of cardenolide concentrations in three samples of overwintering monarch butterflies from Site Alpha in Mexico, January 1980. The oriole and grosbeak samples are partly eaten butterflies dropped to the ground by the birds and collected by us during their feeding bouts in the butterfly colony. The random sample consists of undamaged live butterflies collected within the colony. Two wings per butterfly were analysed. To verify that wing concentration reflects that of the body, we separately analysed the wings and intact bodies of 50 male and 50 female monarch butterflies and found a high correlation between μg of cardenolide per 0.1 g of wing and the total μg per wingless body; $n = 100$, $r = 0.88$, $F = 320$, $P < 0.0001$. The figure shows that the concentrations of the cardenolides in the butterflies damaged by both bird species are distributed similarly to those of the random collection. Because the three distributions are skewed, we subjected all data to $\log_{10}$ transformations and compared the three means by one-way analysis of variance ($F_{2,466} = 0.91$, $P = 0.40$). These data provide strong evidence that cardenolides give individual butterflies at Site Alpha no obvious protection from either orioles or grosbeaks.

some of the overwintering butterflies which were rejected by the birds without being killed may be protected by other distasteful or toxic substances, such as pyrrolizidine alkaloids^{27,28}. Alternatively, the chemical defence of the monarch butterfly may be ineffective in these overwintering sites and the only advantage to individuals within the colonies may derive solely from a reduced probability of being killed, comparable with the advantages of herding^{1,29}. It is possible that the extensive predation is a recent phenomenon attributable to changes in the relative abundances of the milkweed species in eastern North America.

This research was supported by NSF grant DEB78-10658 with L.P.B. as principal investigator, and by the Leslie T. Webster Fund of Amherst College. We thank E. Romero, J. and E. de Jesus, R. Barthelemy, L. Liscomb, B. Olds, S. Friedrich and R. Waide for help in Mexico, R. Lindley for liaison, S. Glazier for laboratory analyses, P. Walford and B. Steele for computing assistance, J. Lane for the Santa Cruz sample, A. Krause and B. Tiffany for technical help, and K. Anderson and T. Lloyd-Evans of Manomet Bird Observatory and M. Clench and R. Leberman of Powdermill Nature Reserve for orioles and grosbeaks. We also thank K. Brown, W. Calvert, F. Chew, C. Nelson, T. Sargent, J. Seiber and N. Stamp for comments on the manuscript, and K. D. Sargent, M. Fenner and E. Belcher for artwork.

Received 10 November 1980; accepted 19 March 1981.

1. Calvert, W. H., Hedrick, L. E. & Brower, L. P. *Science* **204**, 847–851 (1979).
2. Reichstein, T., von Euw, J., Parsons, J. A. & Rothschild, M. *Science* **161**, 861–866 (1968).
3. Brower, L. P., Ryerson, W. N., Coppinger, L. L. & Glazier, S. C. *Science* **161**, 1349–1351 (1968).
4. Swynnerton, C. F. M. *Afr. orn. Un.* **11**, 32–108 (1915).
5. Parsons, J. A. *J. Physiol., Lond.* **178**, 290–304 (1965).
6. Dixon, C. A., Erickson, J. M., Kellett, D. N. & Rothschild, M. *J. zool Res.* **185**, 437–467 (1978).
7. Brower, L. P. in *Biochemical Coevolution* (ed. Chambers, K. L.) 69–82 (Oregon State University Press, Corvallis, 1970).
8. Rothschild, M. & Kellett, D. N. *J. Ent.* **46**, 103–110 (1972).
9. Brower, L. P., Fink, L. S. & Waide, R. B. (in preparation).

Table 2 Cardenolide concentrations (μg per 0.1 g dry weight of the wings) of monarch butterflies eaten to various degrees by black-headed grosbeaks at Site Alpha, January 1980

Damage category and description	Sample size	Cardenolide concentration (μg per 0.1 g dry weight)		
		$\bar{x}$	s.d.	Range
a Lacking abdomen and 2–3 wings	41	40	41	0–128
b Lacking abdomen and 1 wing	19	42	42	0–129
c Lacking abdomen only	30	42	47	0–207
d Abdominal contents squeezed out, exoskeleton not eaten	9	20	26	0–77
e Abdomen slightly squeezed	7	84	106	0–290
f Thorax damage only	18	46	50	0–199
g Lacking head only	10	78	65	0–186
h Lacking 1–3 wings only	12	35	56	0–177
i Moribund (unable to fly but with no visible damage)	31	33	51	0–169
Entire sample	177	42	51	0–290

The butterflies are grouped into nine damage categories, the mean cardenolide concentrations of which do not differ significantly (one way ANOVA $F_{8,176} = 1.17$, $P = 0.32$). The absence of an obvious relationship between degree or kind of damage and cardenolide concentration indicates that once they have attacked monarchs, grosbeaks, unlike orioles, feed indiscriminately on them.

10. Chaney, S. G. & Kare, M. R. *J. Am. vet. med. Ass.* **149**, 938–943 (1966).
11. Benson, J. N. *et al. Toxicol.* **17**, 155–165 (1979).
12. Brower, L. P., McEvoy, P. B., Williamson, K. L. & Flannery, M. A. *Science* **177**, 426–429 (1972).
13. Brower, L. P., Edmunds, M. & Moffitt, C. M. *J. Ent.* **49**, 183–196 (1975).
14. Brower, L. P. & Moffitt, C. M. *Nature* **249**, 280–283 (1974).
15. Brower, L. P. *Scient. Am.* **220**, 22–29 (1969).
16. Brower, L. P., Seiber, J. N., Nelson, C. N. & Bond, C. (in preparation).
17. Roeske, C. N., Seiber, J. N., Brower, L. P. & Moffitt, C. M. *Recent Adv. Phytochem.* **10**, 93–167 (1976).
18. West, D. A. *Auk* **79**, 399–424 (1966).
19. Brower, L. P. & Glazier, S. C. *Science* **188**, 19–25 (1975).
20. Brower, L. P., Brower, J. V. Z. & Corvino, J. M. *Proc. natn. Acad. Sci. U.S.A.* **57**, 893–898 (1967).
21. Tuskes, P. M. & Brower, L. P. *Ecol. Ent.* **3**, 141–153 (1978).
22. Brower, L. P., Calvert, W. H., Glazier, S. C. & Sheppard, M. (in preparation).
23. Urquhart, F. A. & Urquhart, N. R. *Can. J. Zool.* **56**, 1759–1764 (1978).
24. Woodson, R. E. Jr. *Ann. M. bot. Gdn* **41**, 1–211 (1954).
25. Brower, L. P., Seiber, J. N., Nelson, C. N., Lynch, S. P. & Tuskes, P. M. (in preparation).
26. Urquhart, F. A. & Urquhart, N. R. *Can. J. Zool.* **93**, 41–47 (1979).
27. Rothschild, M. & Marsh, N. *Ent. exp. appl.* **24**, 437–450 (1978).
28. Edgar, J. A., Boppre, M. & Schneider, D. *Experientia* **35**, 1447–1448 (1979).
29. Edmunds, M. *Defence in Animals* (Longman, New York, 1974).

Eclosion hormone may control all ecdyses in insects

James W. Truman*, Paul H. Taghert,
Philip F. Copenhaver, Nathan J. Tublitz
& Lawrence M. Schwartz

Department of Zoology, University of Washington, Seattle,
Washington 98195, USA

The moulting process in insects is initiated and directed by the hormone 20-hydroxyecdysone¹, which causes the secretion of a new cuticle and the internal changes that are necessary for the new stage. When the new cuticle is complete, the old one is then shed or ecdysed. In Lepidoptera a neurosecretory peptide, eclosion hormone (EH)^{2–4}, triggers adult ecdysis by acting on the central nervous system to elicit the ecdysial motor programmes⁵ and also causes other physiological changes associated with ecdysis⁶. Although it was originally thought that EH was used only for the rather specialized adult ecdysis³, we report here that the hormone is used for all postembryonic ecdyses in the life history of moths, but its source varies with developmental stage. Moreover, EH may be common to a wide variety of insects. For the ecdyses of the adult⁴, pupa⁷ and the 5th instar larva of the tobacco hornworm, *Manduca sexta* (P.F.C., unpublished results), EH appeared in the blood before the start of the behaviour. Also, injections of partially purified EH caused precocious ecdysis of the three stages with 50% responses occurring at 0.3, 0.02 and 0.01 EH units, respectively.

Although EH seems to be associated with the ecdyses of the adult, pupa and 5th instar larva, the source of the blood-borne hormone varies. EH is distributed in the CNS between the brain and its neurohaemal organ, the corpora cardiaca (CC), and the ventral chain of segmental ganglia⁸. The roles of these two regions in the control of ecdysis were examined by determining their EH titres throughout the life history of *Manduca*.

As seen in Fig. 1, $\sim 10^{-3}$ units of EH activity were extracted from heads or ventral ganglia at the time of hatching. Through the five larval instars, both tissues showed accumulation of hormone during the feeding period of each stage, followed by a plateau during the time of formation of the new cuticle for the next stage. In the case of the 5th instar, this plateau extended from the start of the wandering stage through the period of pupal cuticle production. At each ecdysis, a sharp drop was then seen in the hormone stores of the ventral ganglia. A drop was also observed in the hormone content of the head during ecdysis of the 2nd instar but no further depletions were detected in either whole heads or brains at subsequent larval ecdyses or at pupal ecdysis. Examination of the CC before pupal ecdysis revealed no detectable EH activity.

Adult development was marked by a gradual increase in the titre of hormone in the ventral ganglia. No depletion was seen at adult ecdysis and accumulation continued into adult life. Titre changes in the brain and CC during this time were more complex. Hormone levels increased in the brain through the first half of development but gradually declined during the second half. These changes were accompanied by a >100 -fold increase in activity in the CC. This accumulation of hormone in the CC was apparently due to hormone transferred from the brain because when brains were removed from animals early in adult development and their CC later assayed for the presence of EH, the latter lacked any activity ($n = 5$). At adult eclosion the CC showed a 70% loss of activity. EH activity persisted in both structures through at least the first 4 days of adult life.

A major feature of the titre curves in Fig. 1 is the prominent depletion in stored hormone which occurs at each ecdysis. At the ecdyses of the adult, pupa and 5th instar larva, each drop coincides with the appearance of EH activity in the blood. Moreover, the magnitude of the depletions (0.7, 0.06 and 0.05 EH units, respectively) compares favourably with the dosages required to provoke ecdysis in the same stages. On the basis of these data we have concluded that the sudden drops observed at each of the six ecdyses represent a release of EH into the blood and therefore that all ecdyses in *Manduca* are triggered by EH.

A role for EH during embryonic development was determined using the relatively large embryos of the moth *Hyalophora cecropia*. Preliminary experiments indicated that EH was present in embryos, mainly in the head region. As seen in Fig. 2, significant quantities of EH activity appeared in the head on day 5 of development. The titre remained high for the next 2 days but then dropped sharply between days 7 and 8 and remained low until the time of hatching on day 10. This drop in stored hormone coincided with the ecdysis of an embryonic cuticle by the developing 1st instar larva within the egg. Thus, it seems that EH is involved in embryonic as well as postembryonic ecdyses.

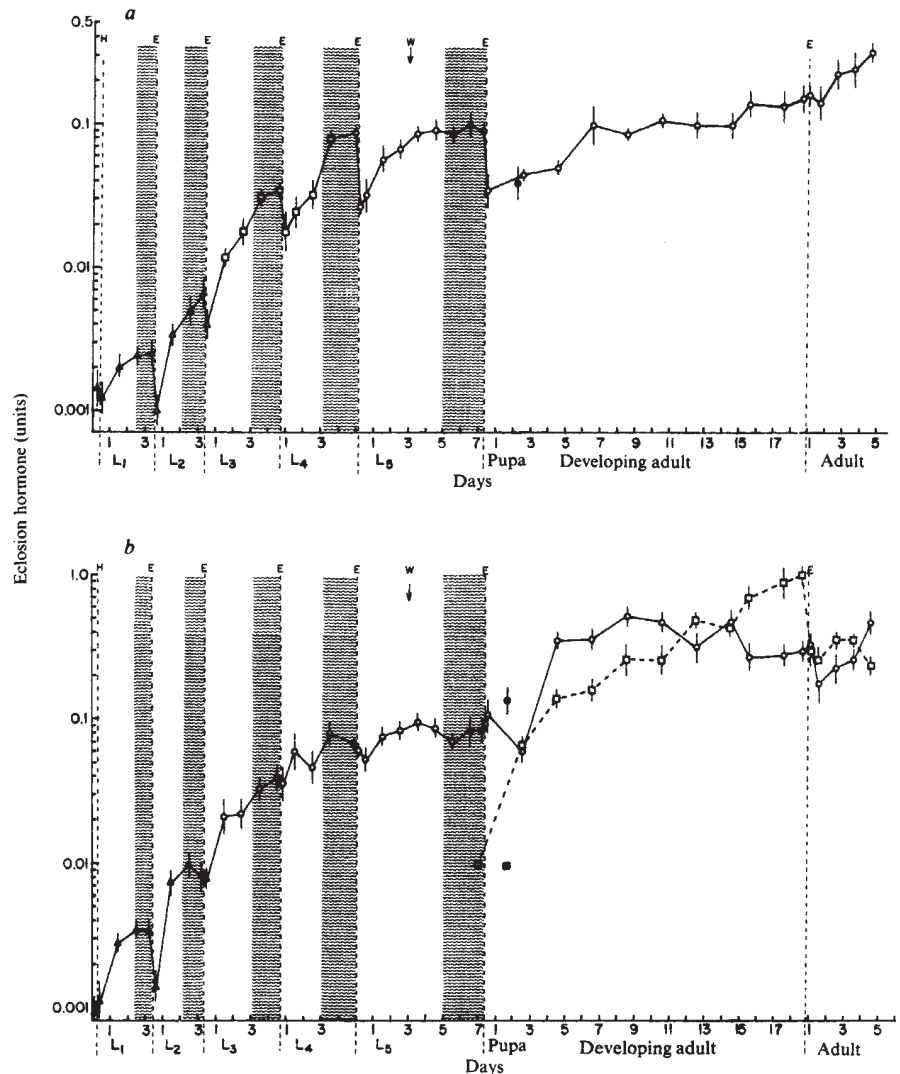
Our data suggest that EH storage and release occurred in either of two regions of the nervous system—the brain–CC complex or the ventral ganglia. Partial chemical characterization of the ecdysis stimulating activity from both regions revealed a single peptide with properties extremely similar or identical to EH⁸. The two regions seem to function independently. Indeed, experiments on adult ecdysis of the giant silkworms show that the brain controls EH release at this stage², whereas at an earlier stage (pupal ecdysis) in *Manduca*, the previous removal of the brain from wandering stage larvae did not interfere with the subsequent release of EH from the ventral ganglia⁸. These experiments, albeit on different moth species, together with the data in Fig. 1, suggest two anatomically distinct systems for the synthesis and release of EH. However, the final proof requires the identification of the EH neurones in both regions.

Release occurs from the ventral ganglia during all postembryonic ecdyses, except for that of the adult which is controlled by the brain–CC axis. In these ecdyses, the choice of release site is correlated with the nature of the control over EH release. The time of adult ecdysis is controlled by a circadian clock which, at least in the case of the silkworms², is in the brain. This clock regulates adult ecdysis through the brain–CC system³. By contrast, there is no direct circadian control over the larval and pupal ecdyses, but they occur at a developmentally fixed time after the start of the moult⁹. These developmentally triggered ecdyses are controlled by the ventral ganglia neuroendocrine system. A circadian involvement in the embryonic ecdysis has not been determined. Thus it is not known whether the above correlation also holds in the embryo. The use of the brain system by the embryo may also be due to its being the earliest part of the CNS to mature¹⁰. The transition from brain regulation in the embryo to regulation by the ventral ganglia in larvae seems to occur gradually as ecdysis to the 2nd-stage larva uses both systems.

The search for EH in insects other than Lepidoptera has been suggestive but not conclusive^{11,12}. However, in these studies

* To whom correspondence should be addressed.

Fig. 1 Eclosion hormone activity extracted from the nervous system of *M. sexta* during its larval, pupal and adult stages. Depending on the stage, tissues from 5 to 300 animals were pooled, homogenized in saline with phenylthiourea, heat treated at 80 °C, centrifuged and the supernatant assayed for EH activity by injection into staged *Manduca* prepupae. The EH activity is expressed in units per animal (one unit being defined as the activity present in a CC pair from a *Manduca* on the last day of adult development). *a*, EH activity stored in the ventral ganglia. Symbols represent the hormone present in: Δ , mid-ventral strip of integument that included the segmental ganglia; $\square$, abdominal and last two thoracic ganglia; $\circ$, abdominal ganglia. *b*, EH activity present in cephalic structures. Symbols represent activity from extracts of: Δ , head and prothorax (the latter taken to insure inclusion of the CC); $\circ$, brain and $\square$, CC pair (the first two values represent the maximum amount of activity that could have been present but not detected in the conditions of the assays). At the times of hatching (H) and ecdyses (E), animals were selected ~4 h before (as judged by developmental markers) and 1 h after the event. W, start of wandering stage. The shaded areas represent the approximate time that new cuticle is being made in preparation for ecdysis (not indicated for adult development). The time marks along the abscissa correspond to lights-off for each day. Each determination is the mean ($\pm$ s.e.) of at least nine assays. Solid symbols are from diapausing pupae.



only the brain was examined as a possible hormone source. The finding that the ventral ganglia controlled most of the ecdyses in *Manduca* prompted us to examine the ventral ganglia as well as the brain in a number of species. A preliminary survey revealed EH activity (as evidenced by the stimulation of precocious pupal ecdysis in *Manduca*⁷) in heat-treated extracts of nervous systems of juvenile forms from the hemimetabolous orders, Orthoptera (*Schistocerca*, *Melanopus*, *Periplaneta*) and Hemip-

tera (*Pyrrhocoris*), and the holometabolous orders, Coleoptera (*Dytiscus*), Trichoptera (*Dicosmoecus*) and Diptera (*Tipula*, *Calliphora*). These preliminary results should be regarded with caution because we have not yet shown that these activities from other orders are chemically similar to the moth EH, and that they are used to stimulate ecdysis in the insects in which they are found. Nevertheless, these results suggest that EH is present in a wide variety of insects and that it provides a general mechanism by which ecdyses are triggered in insects.

We thank Dr Shirley Reiss for technical assistance during part of this study and Professor Lynn M. Riddiford for helpful comments on the manuscript. The work was supported by NSF grant PCM 77-24878 and NIH grant R01 NS13079. P.H.T. and N.J.T. were supported by NSF predoctoral fellowships, L.M.S. by a NIH training program and J.W.T. by a NIH career development award.

Received 23 December 1980; accepted 3 March 1981.

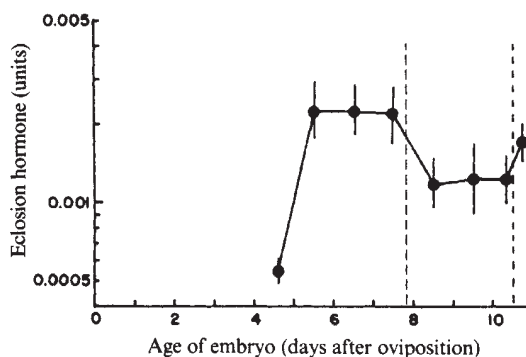


Fig. 2 Eclosion hormone activity extracted from the head and prothoracic segment of *Hyalophora cecropia* embryos during embryonic development. Activity is expressed as units per animal. Symbols represent the mean ($\pm$ s.e.) of eight determinations. The time marks correspond to lights-off. The dashed vertical lines indicate the times of ecdysis of the embryonic cuticle and of hatching, respectively.

1. Riddiford, L. M. & Truman, J. W. in *Insect Biochemistry* (ed. Rockstein, M.) 307-357 (Academic, New York, 1978).
2. Truman, J. W. & Riddiford, L. M. *Science* **167**, 1624-1626 (1970).
3. Truman, J. W. *Biol. Bull. mar. biol. Lab., Woods Hole* **144**, 200-211 (1973).
4. Reynolds, S. E., Taghert, P. H. & Truman, J. W. *J. exp. Biol.* **78**, 77-86 (1979).
5. Truman, J. W. *J. exp. Biol.* **74**, 151-174 (1978).
6. Truman, J. W. in *Insect Biology in the Future* (eds Locke, M. & Smith, D. S.) 385-401 (Academic, New York, 1980).
7. Truman, J. W., Taghert, P. H. & Reynolds, S. E. *J. exp. Biol.* **88**, 327-337 (1980).
8. Taghert, P. H., Truman, J. W. & Reynolds, S. E. *J. exp. Biol.* **88**, 339-349 (1980).
9. Truman, J. W. *J. exp. Biol.* **57**, 805-820 (1972).
10. Khan, T. R. & Fraser, A. in *Neurosecretion* (eds Heller, H. & Clark, R. B.) 349-369 (Academic, London, 1962).
11. Carlson, J. R. *J. comp. Physiol.* **115**, 299-317 (1977).
12. Miller, P. L. & Mills, P. S. in *Perspectives in Experimental Biology* Vol. I. (ed. Spencer-Davis, P.) 199-208 (Pergamon, Oxford, 1976).

Positional signal transmission in the developing chick limb

Lawrence S. Honig*

Department of Biology as Applied to Medicine, Middlesex Hospital Medical School, London W1P 7PN, UK, and Division of Developmental Biology, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

* Present address: Laboratory of Developmental Biology, University of Southern California, GER-319, PO Box 77912, Los Angeles, California 90007, USA.

It has been suggested that the pattern along the antero-posterior axis of the embryonic developing chick limb arises from a gradient of diffusible morphogen^{1,2} produced by a special region of cells on the posterior margin of the limb bud, called the polarizing region³. Grafts of posterior polarizing-region tissue to an anterior site in an embryonic chick wing bud result in mirror-image duplication of the limb bud; typically three extra wing digits are produced (Fig. 1f, g). Direct evidence for a long-range signal has, however, been lacking, and recently it has been suggested⁴ that chick wing development might better be understood as resulting only from local, neighbour-neighbour-like interactions of the sort postulated to occur during regeneration of insect and amphibian limbs and insect imaginal disks⁵. As we report here, operations were performed in which chick wing buds were made host to grafts of leg bud responding tissue adjacent (posteriorly) to a grafted polarizing region (Fig. 2). The results showed that the pattern signal from the polarizing region can be transmitted through the leg tissue over distances of tens of cell diameters.

Double grafts in which leg tissue provides a living 'barrier' between grafted wing polarizing region and distal host chick wing tissue, are shown diagrammatically in Fig. 2. Leg tissue maintains its character in wings and when placed at wing bud tips develops into toes⁶ which are recognizable through the patterns of their distinctively shaped cartilage elements and associated tissue (Fig. 1h and legend). (Many grafts also used quail marked tissue and the details of these will be presented elsewhere.) Sixty-six grafts were performed as described in Fig. 2 using leg anterior pieces ranging in width from 100 to 350 μm . In nearly all (61) cases the leg tissue, as expected, developed into recognizable toes. The important question was whether the signal from a grafted polarizing region could influence the wing tissue on the posterior side of the leg tissue barrier. In 34 cases (average width $215 \pm 70 \mu\text{m}$ s.d.), only host wing digits were found posteriorly to toes (Fig. 1e), but in another 27 cases (average width $189 \pm 57 \mu\text{m}$ s.d.) mirror-image duplicated wing digits appeared posterior to the toes (Fig. 1a-d). In general, leg anterior grafts of narrower width tended to yield limbs of this latter category, in which the posterior host tissue was influenced, while anterior grafts of large width provided more examples of posterior host tissue being unaffected.

Interpretations of these experiments depend on several considerations. First, control grafts of anterior leg tissue, without the additional polarizing region grafts, commonly did not give rise to any extra digits (Fig. 1g) in the resultant chick wings: out of 32 such operations, only 5 wings presented extra cartilage elements of which 4 possessed toe digits I and 1 had only an unidentifiable cartilage nodule. Therefore, it was the long-distance influence of the grafted polarizing region, not the interaction between the graft of the leg anterior tissue and the

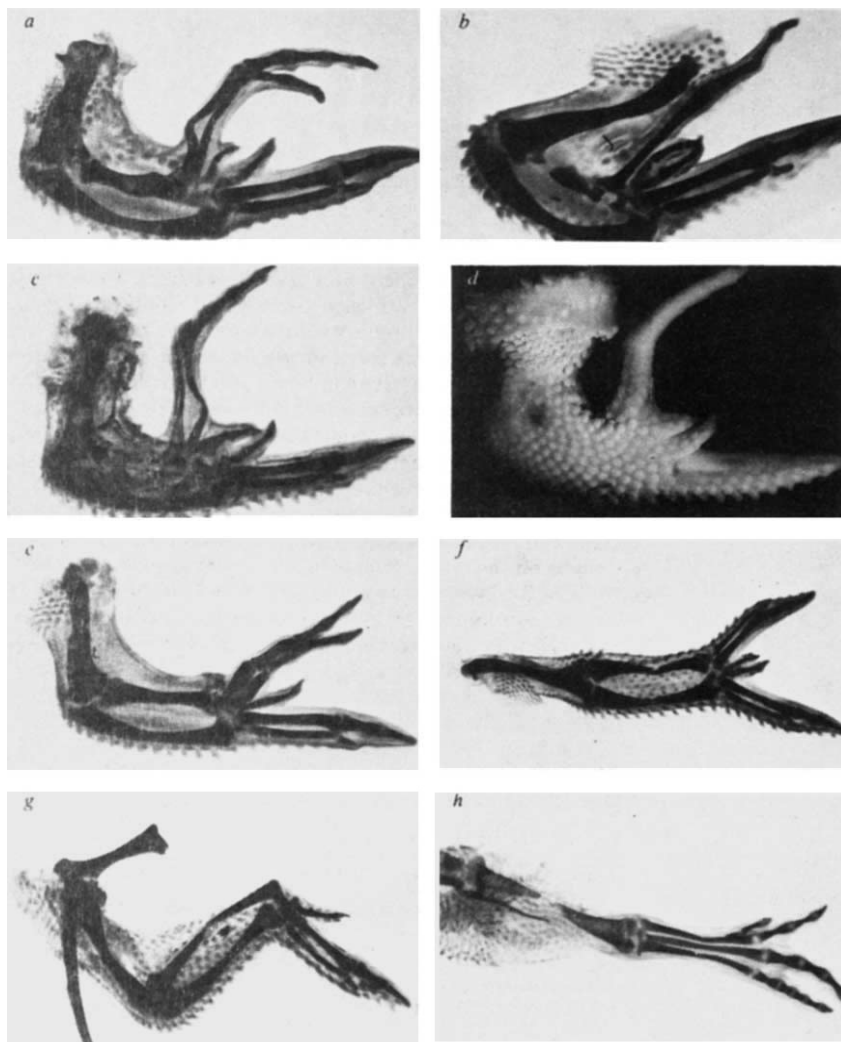


Fig. 1 Wholemount limbs from 10-12-day chick embryos (magnifications about $\times 6$) following fixation, staining with alcian green 2GX, dehydration with ethanol and clearing with methyl salicylate⁹; limb d was not cleared. Wing digits are denoted by Arabic numerals; leg digits are represented by Roman numerals. Digit patterns are specified in anterior to posterior order. Toe digits (h) differ from wing digits (g) by the long thin metatarsals of II, III and IV and their more numerous phalanges. Toe I has some similarity to wing digit 2 but is distinguished by several features. Toe I normally arises not at the wrist but as an incomplete metatarsal of inverted triangular shape at distal metatarsal level. By contrast, metacarpal 2 arises at the wrist and is of thick rectangular shape with an anterior 'knob' at midlength. The first phalanx of toe I is, in common with other proximal toe phalanges, of hour-glass shape: the basic cylindrical shape is constricted at the centre so as to resemble an 'X'; the first phalanx of digit 2 is more like a narrow truncated cone whose proximal width is greater than that distally and is not constricted in the centre. The soft tissue of toe digits clings closely to the skeleton whereas it is more muscular and filled out for wing digits. At late stages, feather germs are present on avian wing digits but are not present on the toes which only develop distinctive scale primordia in these chicks and quail. a-d, Examples in which posterior host wing was influenced through a leg barrier: digit patterns are (a) III II 2 2 3 4, (b) II 2 2 3 4 and (c) 4 III 2 2 3 4. In d, specimen c is shown before it was cleared, showing the feather germs on the duplicate host digit (digit 2) and scales on the leg digit. e, Result of an operation in which posterior host wing was unaffected; the identifiable digit pattern is III II 2 3 4. f, Wing reduplication following graft of polarizing region only, with digit pattern 4 3 2 2 3 4. g, Wing with normal set of cartilage elements found after graft of leg anterior tissue alone, with digit pattern 2 3 4. h, Distal elements (tibia-fibula and foot) of normal control unoperated leg showing digit pattern I II III IV.

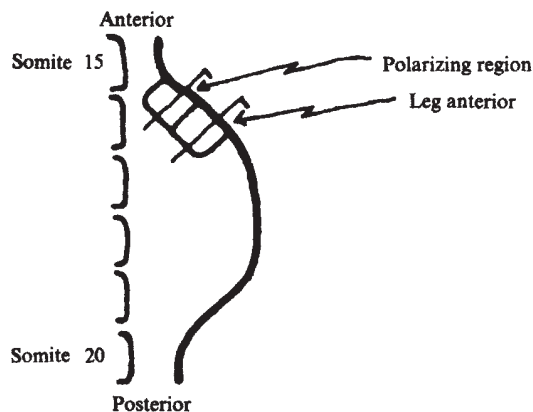


Fig. 2 Diagram of operation consisting of double graft, of wing polarizing region (donor stage 20–22) and leg anterior tissue (donor stage 19–23), each pinned with a platinum pin made from 25- μ m diameter wire, replacing host wing bud tissue opposite somites 16/17. Hosts were chick embryos stage 19–21 and received several drops of Hank's balanced salts solution containing 20 mM HEPES and 1% antibiotic-antimycotic (Gibco) before being resealed with Sellotape and reincubated at 38 °C.

host, which caused the extra wing digits 2 obtained above, following the double grafts. Second, it is possible that polarizing region cells were physically migrating (in a posterior and distal direction) around the leg anterior tissue so as to contact directly the host tissue giving rise to the digits 2. This seems unlikely in the absence of any evidence for such migration in chick wing buds⁷. Nevertheless, control experiments were performed as in Fig. 2 but using anterior tissue from proximal parts of older ($\geq$ stage 22) limb buds. Such tissue has less competence to respond to or transmit the signal of the polarizing region and after such grafting operations 13 cases had no toes at all and 4 had a toe adjacent to host structures; in none of these operations were host structures duplicated. Third, although the host origin of duplicate wing digits 2 of the sort in Fig. 1a–d was indicated by their morphology and feather germs, it was confirmed by histological sections of limbs of this type resulting from operations using quail anterior tissue barriers. The duplicate digits 2 consisted almost entirely (> 90%) of host chick cells.

The ability of polarizing region tissue to influence distal wing tissue a distance away from it is in direct distinction to expectations of local intercalation type models^{4,5}. The experiments here show that the polarizing region can signal patterning information over a distance of about 200 μ m, or about 20 cell diameters. But, there is no evidence for influencing positional information of cells at larger distances of over 300 μ m. The results of several lines of investigation suggest that this distance is, at the stages operated on, roughly the antero-posterior extent of the presumptive limb field⁸; these include the absolute number of marked digits obtained using the double grafts described here (L.H., in preparation), fate maps⁷, barrier experiments interpretable as fate maps⁹, and grafts of two polarizing regions^{10,11}. The long-range communication over ~20 cell diameters may be through a freely diffusible morphogen, but equally might occur through cell-to-cell transferred molecule(s), electric potentials, or even through a sequence of cell surface-surface interactions¹².

I thank Dr Lewis Wolpert for discussion and advice, and acknowledge the support of the Anna Fuller Fund (Fellowship No. 487) and the MRC.

Received 28 November 1980; accepted 20 February 1981.

1. Wolpert, L. *Curr. Topics dev. Biol.* **6**, 183–224 (1971).
2. Tickle, C., Summerbell, D. & Wolpert, L. *Nature* **254**, 199–202 (1975).
3. MacCabe, A.B., Gasseling, M.T. & Saunders, J.W. Jr *Mechanisms Ageing Dev.* **2**, 1–12 (1973).
4. Iten, L.E. & Murphy, D.J. *Dev. Biol.* **75**, 373–385 (1980).
5. French, V., Bryant, P.J. & Bryant, S.V. *Science* **193**, 969–981 (1976).

6. Saunders, J.W. Jr, Gasseling, M.T. & Cairns, J.M. *Dev. Biol.* **1**, 281–301 (1959).
7. Stark, R.J. & Searls, R.L. *Dev. Biol.* **33**, 138–153 (1973).
8. Summerbell, D. & Honig, L.S. *Am. Zool.* **21** (in the press).
9. Summerbell, D. *J. Embryol. exp. Morph.* **50**, 217–233 (1979).
10. Wolpert, L. & Hornbruch, A. *J. Embryol. exp. Morph.* (in the press).
11. Summerbell, D. *J. Embryol. exp. Morph.* (in the press).
12. Babloyantz, A. *J. theor. Biol.* **68**, 551–561 (1977).

Neurotensin blocks certain amphetamine-induced behaviours

Gregory N. Ervin, Larry S. Birkemo,
Charles B. Nemeroff & Arthur J. Prange Jr

Biological Sciences Research Center, Departments of Psychiatry, Psychology and the Neurobiology Program, University of North Carolina, School of Medicine, Chapel Hill, North Carolina 27514, USA

Bilateral injections of either neurotensin (NT; 0.3, 1 or 5 μ g in 1 μ l artificial CSF) or haloperidol (HA; 2.5 or 5 μ g in 1 μ l 0.3% tartaric acid) into nucleus accumbens of rats markedly diminished the forward locomotion and rearing induced by d-amphetamine (AM; 2 mg per kg, IP). Neither NT nor HA affected the insomnia or sniffing component of AM arousal. Isovolumetric intra-accumbens injections of artificial CSF or the endogenous decapeptide, luteinizing hormone-releasing hormone (LHRH; 3 μ g), did not affect AM behaviours. Since intra-accumbens injections of NT (1 μ g) or HA (2.5 μ g) neither altered forward locomotion or rearing observed in untreated rats placed in an open field nor a variety of reflex activities, the observed effects of NT and HA in AM-treated rats were probably not due to impaired motor function *per se*. In contrast, NT does not produce neuroleptic-like effects when injected into nucleus caudatus; HA (5 μ g) blocked stereotyped sniffing, licking, biting and head bobbing observed after AM (5 mg per kg, IP), but NT (3 or 5 μ g) did not. Since NT and dopamine are present in substantial quantities in the nucleus accumbens, NT may act in the nucleus accumbens to modulate dopaminergic function.

Neurotensin (NT), a tridecapeptide (pGlu-Leu-Tyr-Glu-Asn-Lys-Pro-Arg-Arg-Pro-Tyr-Ile-Leu-OH), is an endogenous brain peptide which satisfies several neurotransmitter criteria^{1,2} and may participate in both endocrine and behavioural functions^{3,4}. Radioimmunoassay and immunohisto-fluorescence studies of rat brains indicate that certain diencephalic regions may have especially high concentrations of NT⁵. The nucleus accumbens is rich in both NT and dopamine (DA). Nucleus accumbens DA is contained in a rich plexus of nerve terminals that is an important part of the forebrain DA terminal system, which functions in arousal and exploration^{6,7} and mediates increased forward locomotion and rearing stimulated by low doses of (+)amphetamine (AM)^{8,9}. We describe here experiments which show that NT and the neuroleptic haloperidol (HA) have similar effects on AM-induced behaviour when injected into the nucleus accumbens of the rat, but that NT differs from HA in having no effect on AM behaviour when injected into the caudate. This suggests that NT may have a neuroleptic action without some of the side effects of drugs such as HA.

To test for possible interactions of NT and DA systems within the nucleus accumbens, we observed the effects of bilateral intra-accumbens microinjections of NT on the behavioural responses produced by a fixed dose of AM, and which are believed to be mainly dependent on DA release. Using a standardized technique of time sampling and behavioural scoring⁹, we observed that NT attenuates AM-induced locomotion and rearing but not the associated insomnia or sniffing. Similar effects were also produced by bilateral intra-accumbens injections of HA, a DA receptor antagonist. We also studied the interactions of NT and the striatal DA system by comparing the effects of NT and HA on behavioural response to a high dose of

Table 1 Effects of injections into nucleus accumbens on selected behavioural responses to amphetamine

	CSF	Neurotensin (μg)				LHRH (3 μg)	Tartaric acid	Haloperidol (μg)	
		0.1	0.3	1	5			2.5	5
Behavioural items (n)	30	10	9	11	15	8	10	10	10
Forward locomotion	15 (1-36)	8.5 (2-28)	6* (0-28)	8* (0-20)	5+ (0-21)	14.5 (1-30)	14 (2-33)	3+ (0-8)	1‡ (0-30)
Rearing	4.5 (0-38)	2 (0-26)	1* (0-14)	0* (0-11)	0‡ (0-11)	4.5 (0-13)	1 (0-4)	1 (0-3)	0 (0-11)
Sniffing	39.5 (10-53)	44 (18-50)	35 (27-40)	28 (11-46)	34 (12-54)	38.5 (19-50)	40 (30-51)	39.5 (26-48)	25* (0-49)
Sleeping	2.5 (0-18)	5 (0-15)	3 (0-15)	6 (0-18)	0 (0-15)	0 (0-12)	0 (0-12)	3 (0-21)	7.5 (0-22)

Medians and ranges of behavioural scores over 3 h after (+)amphetamine (2 mg per kg i.p.) are shown. Immediately before amphetamine administration, all rats received injections into nucleus accumbens through indwelling cannulae. Doses shown were delivered bilaterally in 1 μl injectate over 1 min. Animals were observed for 30 s every 10 min for 3 h, and scored by a 'blind' observer using a standard sampling technique. The maximum possible score for each item was 54. Peptide scores were compared with artificial CSF scores, and HA scores with tartaric acid scores, by the Mann-Whitney *U*-test.

* $P < 0.05$. + $P < 0.001$. ‡ $P < 0.01$.

AM acting on the striatum^{8,10}. Although intracaudate injections of HA blocked AM-induced stereotyped behaviour, injections of NT did not.

Male Sprague-Dawley rats (350–450 g, Zivic Miller) were housed in individual cages at a temperature of $\sim 23^\circ\text{C}$ with 12 h of light (0600–1800 h). Rats were cannulated under ketamine and phenobarbital anaesthesia when mounted in a stereotaxic instrument with the incisor bar 5 mm above the inter-aural line. Guide cannulae (22-gauge stainless steel), affixed to the cranium with dental acrylic, ended 3 mm dorsal to the sites of interest in the medial nucleus accumbens (9.6 mm anterior to inter-aural line, 1.25 mm right and left of sagittal sinus, and 6 mm ventral to dura) or caudate head (8.6 mm anterior to inter-aural line, 2.5 mm right and left, and 5 mm ventral to dura). These cannulae were implanted at least 2 weeks before behavioural testing. Stainless steel injection cannulae were 30 gauge. After testing, verification of accumbens cannulae placements was made in 70 out of 75 rats by frontally sectioning freshly frozen brains; 132 cannulae tracks terminated in the anteromedial nucleus accumbens, and eight were within 0.5 mm of this structure. Caudate cannulae placements were examined in 25 out of 30 rats; all cannulae tracks terminated within the nucleus caudatus.

On a test day, rats were placed individually in clear Plexiglass test cages (21 $\times$ 44 $\times$ 20 cm high) with wire mesh floors. Each test cage was mounted in a wooden box, with a smoked Plexiglass window (14 $\times$ 65 cm) which allowed viewing of a rat from ~ 20 –40 cm. Test cages were lit with a 7.5 W lamp; the experimental room was dark. After an habituation period of 2 h, rats were removed from test cages and injection cannulae, connected by polyethylene tubing (PE 20) to a 10- μl Hamilton syringe and automatic dispenser, were inserted into nucleus accumbens or nucleus caudatus bilaterally; 1 μl of fluid was injected over 1 min into each site. Each injection contained 0.1, 0.3, 1.0 or 5.0 μg NT (in artificial cerebrospinal fluid (CSF), pH 7.4, Beckman); 3.0 μg luteinizing-hormone releasing hormone (LHRH, in CSF, Parke-Davis); 2.5 or 5.0 μg HA (in 0.3% tartaric acid, pH 3.0, MacNeil Laboratories); or CSF¹¹ or tartaric acid vehicles alone. Immediately after intra-accumbens injections, rats were injected intraperitoneally (i.p.) with 2 mg per kg AM (Smith, Kline and French), returned to test cages, and observed over the next 3 h by an observer ignorant of the treatments. Animals with cannulae in the nucleus caudatus were treated in the same way except that they received a higher dose of AM (5 mg per kg). Occurrence of the following six behavioural items was recorded in the rats with nucleus accumbens cannulae⁶: (1) forward locomotion—movements of the rat from one end of the cage to the other; (2) rearing; (3) sniffing—rapid and repetitive movement of the nostrils and/or snout; (4) sleeping—any resting posture with eyes closed throughout a

whole 10-s interval; (5) rapid bobbing of the head; (6) grooming. In rats with nucleus caudatus cannulae receiving the higher dose of AM, the following behaviour types were considered to be stereotype indices: head bobbing (as described above), biting of the wire mesh floor, licking (of the floor or walls), and sniffing stereotype—continuous rapid and repetitive movement of nostrils and/or snout directed in patterns towards the floor or walls. Behaviour of each rat was observed for 30 s every 10 min for 3 h, and each 30-s observation was divided into three 10-s intervals. When a behaviour type was observed in a 10-s interval, that behaviour received a score of 1. Thus the total score for each behavioural item indicated the number of 10-s intervals in which that behaviour occurred, with a maximum possible score, after injection, of 54. It should be emphasized that these scores represent the incidence of behaviours; except for sniffing or sniffing stereotypes, we did not attempt to rate intensity of behaviours. Scoring of sniffing and sniffing stereotype was mutually exclusive. Using the Mann-Whitney *U*-test¹², we

Table 2 Effects of injections into caudate nucleus on selected behavioural responses to amphetamine

	CSF	Neurotensin (μg)		Tartaric acid	Haloperidol (5.0 μg)
		3.0	5.0		
Behavioural items (n)	13	10	10	12	10
Stereotypes (consisting of sniffing stereotype, head bobbing, biting and licking)	72 (46-98)	59.5 (32-92)	57 (14-98)	67 (45-99)	4.5* (0-39)
Forward locomotion	9 (0-53)	10 (1-33)	7 (0-16)	3 (0-26)	16* (1-52)
Rearing	2 (0-14)	3 (0-32)	3 (0-9)	1 (0-26)	7.5 (0-37)
Sniffing	11 (2-32)	10.5 (2-25)	20.5 (0-47)	13 (4-19)	38.5* (28-47)
Sleeping	0 (0)	0 (0)	0 (0)	0 (0)	0 (0-2)

Medians and ranges of behavioural scores over 3 h after (+)amphetamine (5 mg per kg i.p.) are shown. Immediately before amphetamine administration, all rats received injections into the caudate nucleus through indwelling cannulae. Doses shown were delivered bilaterally in 1 μl injectate over 1 min. Animals were observed for 30 s every 10 min for 3 h, and scored by a 'blind' observer using a standard sampling technique. Peptide scores were compared with artificial CSF scores, and HA with tartaric acid scores, by the Mann-Whitney *U*-test.

* $P < 0.01$.

Table 3 Effects of injections into nucleus accumbens on open field performance

	Uncannulated		Cannulated		
	Uninjected	CSF	Neurotensin (1 µg)	Tartrate	Haloperidol (2.5 µg)
Behavioural items (n)	5	8	8	5	5
Squares traversed	91 (62–215)	101 (51–249)	101 (14–208)	107 (42–367)	112 (21–188)
Rears	20 (13–30)	18 (3–38)	10 (0–28)	27 (7–40)	11 (5–39)

Medians and ranges of behavioural scores of rats observed for 6 min in an open field are shown. On the basis of their performance in the open field on the previous day, cannulated rats were sorted to establish four comparable groups. Cannulated rats received injections into the nucleus accumbens 6 min before being put into the open field. Bilateral injections of NT (1 µg) and HA (2.5 µg) were given in 1 µl of artificial CSF and 0.3% tartaric acid, respectively. Statistical comparisons were made using the Mann–Whitney *U*-test. Rats injected with NT or HA did not differ from vehicle-injected or uncannulated rats.

compared scores from peptide-treated rats with scores from artificial CSF-treated rats, and HA scores with tartaric acid scores. (To determine if intra-accumbens vehicle injections alone affected AM-induced behaviours, we also observed nine cannulated rats which received AM but no intra-accumbens injections. Vehicle-treated animals were identical to non-injected animals except that tartaric acid significantly reduced amphetamine-induced rearing ($P < 0.05$).)

The four behavioural types affected by the low dose of AM in rats with nucleus accumbens cannulae are shown in Table 1. Nucleus accumbens injections of 0.3, 1.0 or 5.0 µg NT significantly attenuated the forward locomotion and rearing elicited by AM, but did not affect sniffing or insomnia. Injections of 2.5 or 5.0 µg HA also significantly decreased AM-induced locomotion, as described by others¹³. Injections of HA did not decrease insomnia, but the higher dose of HA (5.0 µg) significantly attenuated, but did not block, the sniffing component of AM arousal. The effects of NT on AM arousal were apparently not nonspecific attributes of neuropeptides because LHRH did not significantly alter AM-induced behaviours.

Table 2 shows the behavioural effects of the high dose of AM in rats with intracaudate cannulae. After AM administration, rats locomoted until stereotyped behaviours emerged and then prevailed. Caudate injections of HA (5.0 µg) blocked these stereotyped, and rats locomoted and sniffed as did rats given the lower dose of AM (Table 1). Such a shift in behavioural response to AM is also seen in rats with striatal DA denervations⁸. In contrast to the effects of intra-accumbens injections of NT, intracaudate injections of NT did not have the same effect as HA; intracaudate NT injections (3–5 µg) did not block stereotyped behaviours induced by a high dose of AM, but intrastriatal HA (5 µg) did, as described elsewhere¹³.

We performed two evaluations which indicated that the effects of NT and HA on AM-induced locomotion and rearing were not due to generalized hypokinesia or motor impairment caused by intra-accumbens injections of NT or HA alone. Eight cannulated rats were examined for neurological functions daily for 4 days and showed stable behaviour. We observed hopping reaction in each limb, righting response on, or 30 cm above, a surface, vibrissae placing, chin placing, and the ability to right on an inverted metal grid¹⁴. Immediately after final baseline testing, rats were given intra-accumbens injections of 1 µg NT ($n = 3$), 2.5 µg HA ($n = 2$), or 1 µl CSF ($n = 3$). They were re-examined after 3.25 h by an observer ignorant of the treatments. None of the nucleus accumbens injections altered the performance of rats in these neurological tests (data not shown), suggesting no gross motor debilitation by NT or HA.

We observed open field (OF) performance of other intra-accumbens cannulated ($n = 26$) and uncannulated ($n = 5$) rats during 6-min sessions conducted on 2 consecutive days. A clear Plexiglass OF (80 × 80 × 60 cm high) was placed in a wooden box having a smoked Plexiglass top which allowed viewing of rats from ~90 cm. A grey piece of paper was placed under the OF; black lines drawn on the paper divided the floor into 64 squares

10 × 10 cm. The box was illuminated by two 7.5 W lamps; the experimental room was dark. Rats were placed in the OF facing the corner, and the following two behavioural items recorded: squares traversed—number of squares through which the rat moved all four limbs, and rears—vertical extension of forelimbs in any part of the OF. On the basis of OF performance on day 1, cannulated rats were divided into four equivalent subgroups which did not differ from uncannulated rats (data not shown). On day 2, cannulated rats received intra-accumbens injections containing 1 µg NT ($n = 8$), 2.5 µg HA ($n = 5$), 1 µl CSF ($n = 8$) or 1 µl tartaric acid ($n = 5$) 6 min before being placed in the OF. Scores of NT- or HA-treated rats did not differ significantly from scores of vehicle-treated or uncannulated rats (Table 3), which suggests that attenuation of AM-induced locomotion and rearing by intra-accumbens NT and HA was not due to an inability of rats to execute locomotor or rearing movements.

The neuroleptic-like effect of intra-accumbens NT on AM-induced behaviours is not the only evidence for possible NT–DA interactions in the brain. Like neuroleptics, NT delivered into the ventricular system elicits hypothermia^{15,16}, muscle relaxation¹⁷, and potentiation of pentobarbital-induced narcosis¹⁵. Of particular interest is the recent report¹⁸ that treatment of rats with HA produces marked increases in immunoreactive NT levels in the nucleus accumbens, with only a small increase in the striatum. Furthermore, intracerebrally administered NT has recently been shown to elicit an effect in two commonly used screens for neuroleptic agents, electrical self-stimulation and active avoidance responding¹⁹. Taken together these results suggest that NT may possess anti-dopaminergic (neuroleptic-like) properties²⁰. However, because the precise central synaptic sites where NT and neuroleptics produce these various effects are unknown, the principle of central NT–DA interactions is still tentative. In view of the known anatomy of NT and DA systems, this work suggests that exogenous NT produces its effects at sites where endogenous NT is plentiful (for example, nucleus accumbens), and not where it is scarce (for example, nucleus caudatus) although DA may be concentrated at both sites. Further support for an NT–DA interaction in the nucleus accumbens has recently been provided by the finding that NT, applied iontophoretically, exerts a potent inhibitory effect on DA-sensitive neurones in the nucleus accumbens²¹. Because most clinically useful antipsychotic agents act at all DA receptor sites, they produce various neurological side effects presumably mediated by the nigrostriatal DA system. The finding that NT blocks AM-induced behaviours at a mesolimbic DA site but not at the extra-pyramidal DA sites, suggests the possibility of developing new pharmacological agents with antipsychotic efficacy, but without the unwanted side effects.

A preliminary account of certain of these data has appeared elsewhere²⁰. This work was supported by grants from USPHS.

Received 17 November 1980; accepted 28 January 1981.

- Uhl, G. R. & Snyder, S. H. *Life Sci.* **19**, 1827–1832 (1976).
- Uhl, G. R., Goodman, R. R. & Snyder, S. H. *Brain Res.* **167**, 77–91 (1979).
- Carraway, R. & Leeman, S. E. *J. biol. Chem.* **248**, 6851–6854 (1973).
- Bissette, G., Manberg, P., Nemeroff, C. B. & Prange, A. J. Jr *Life Sci.* **23**, 2173–2182 (1978).

5. Kobayashi, R., Brown, M. & Vale, W. *Brain Res.* **126**, 548-564 (1977).
6. Young, R. C., Ervin, G. N. & Smith, G. P. *Pharmac. Biochem. Behav.* **5**, 565 (1976).
7. Fink, J. S. & Smith, G. P. *J. comp. physiol. Psychol.* **93**, 34-65, 66-84 (1979).
8. Kelly, P. H., Seviour, P. W. & Iversen, S. D. *Brain Res.* **94**, 507-522 (1975).
9. Ervin, G. N., Fink, J. S., Young, R. C. & Smith, G. P. *Brain Res.* **132**, 507-520 (1977).
10. Kennedy, L. A. & Zigmond, M. J. *Brain Res.* **168**, 408-413 (1979).
11. Merlis, J. K. *Am. J. Physiol.* **131**, 67-81 (1940).
12. Goldstein, A. *Biostatistics: an Introductory Text* (Macmillan, New York, 1964).
13. Pijnenburg, A. J. J., Honig, W. M. M. & Van Rossum, J. M. *Psychopharmacologia* **41**, 87-95 (1973); **45**, 65-79 (1975).
14. Sechzer, J. A., Ervin, G. N. & Smith, G. P. *Exp. Neurol.* **41**, 723-737 (1973).
15. Nemeroff, C. B. *et al. Brain Res.* **128**, 485-496 (1977).
16. Bissette, G., Nemeroff, C. B., Loosen, P. T., Prange, A. J. Jr & Lipton, M. A. *Nature* **262**, 607-609 (1976).
17. Osbahr, A. J. III, Nemeroff, C. B., Manberg, P. J. & Prange, A. J. Jr *Eur. J. Pharmac.* **54**, 299-302 (1979).
18. Govoni, S., Hong, J. S., Yang, H.-Y. T. & Costa, E. *J. Pharmac. exp. Ther.* **215**, 413-417 (1980).
19. Luttinger, D., Wiggins, R., King, R. A., Nemeroff, C. B. & Prange, A. J. Jr *Neurosci. Abstr.* **6**, 843 (1980).
20. Nemeroff, C. B. *Biol. Psychiat.* **15**, 283-302 (1980).
21. McCarthy, P. S., Walker, R. J., Yajima, H., Kitagawa, K. & Woodruff, G. N. *Gen. Pharmac.* **10**, 331-333 (1979).

Selective effects of somatostatin-14, -25 and -28 on *in vitro* insulin and glucagon secretion

Lawrence Mandarino, Debra Stenner,
Wayne Blanchard, Steven Nissen & John Gerich*

Endocrine Research Unit, Departments of Medicine and Physiology,
Mayo Medical School and Mayo Clinic, Rochester,
Minnesota 55901, USA

Nicholas Ling, Paul Brazeau, Peter Bohlen,
Frederick Esch & Roger Guillemin

Laboratories for Neuroendocrinology, Salk Institute for Biological
Studies, San Diego, California 92138, USA

The widely occurring¹⁻³ tetradecapeptide somatostatin (SRIF-14) has been variously implicated⁴⁻⁸ as a neurotransmitter, a neurohormone, a cybernin (local regulatory factor) and a hormone. In the first isolation of SRIF-14 from hypothalamic extracts^{9,10} and subsequent extracts of other tissues¹¹⁻¹⁵, peptides of higher molecular weight but with similar activity have been noted. Recently two such peptides have been characterized as the 28-amino acid SRIF-28 (from porcine gastrointestinal tract¹⁶ and porcine and ovine hypothalamus^{17,18}) and the 25-amino acid SRIF-25 (from ovine hypothalamus¹⁷), each of which consists of an N-terminal extension of SRIF-14. We now report that SRIF-28 and SRIF-25 are more potent than SRIF-14 in the inhibition of insulin release, but that SRIF-14 preferentially inhibits glucagon release. This suggests that SRIF-28 and SRIF-25 are not mere biosynthetic precursors of SRIF-14 and that their differential release may be physiologically important.

To compare the relative effects of SRIF-28, SRIF-25 and SRIF-14 on secretion of insulin and glucagon *in vitro*, each peptide was synthesized by solid phase methods¹⁹; their primary structure was recharacterized on a Beckman 8900 amino acid sequencer, and purity confirmed by HPLC. The peptides were dissolved in pancreas perfusate²⁰ and infused by side-arm syringe (0.1 ml min⁻¹) starting 10 min before and continuing for 5 min after a 15 min pancreatic stimulation by arginine (10 mM). Each pancreas was exposed to only a single concentration of a peptide.

Pancreases were isolated from overnight fasted Wistar rats (220-280 g) and perfused *in vitro* with a non-recirculating medium (3 ml min⁻¹) as described elsewhere²⁰. Portal venous effluent was collected at 1-min intervals into chilled tubes containing 25 µM EDTA; after completion of experiments, the effluent was aliquoted and frozen until determination of insulin²¹ and glucagon²² contents by radioimmunoassay. Hormone secretion, presented as the total amount of hormone

released during each 15-min stimulation period (see Table 1), was calculated by summation of the amount of the hormone released for each minute of the 15-min stimulation. Basal glucagon (~0.1 ng min⁻¹) and insulin (~1.5 ng min⁻¹) secretory rates during individual experiments were not significantly different.

SRIF-25 and SRIF-28 did not differ in their biological effects. At all concentrations studied, both peptides inhibited secretion of insulin more potently than did SRIF-14. Linear regression analysis indicated that 50% inhibition of insulin occurred at ~1.4 × 10⁻¹¹ M for SRIF-28 and SRIF-25 and at ~1.5 × 10⁻¹⁰ M for SRIF-14. In contrast, SRIF-14 inhibited secretion of glucagon more potently than did SRIF-25 and SRIF-28; 50% inhibition of glucagon release occurred at ~3.3 × 10⁻¹² M SRIF-14 and at ~3.2 × 10⁻¹¹ M SRIF-25 and SRIF-28. Thus, both SRIF-25 and SRIF-28 seem to inhibit preferentially insulin release whereas SRIF-14 seems to inhibit preferentially glucagon secretion. As a non-recirculating perfusion system was used in these experiments, the different activities observed for these peptides cannot be explained by their differing rates of degradation.

It has been suggested that SRIF-28 may represent prosomatostatin, the biosynthetic precursor of SRIF-14^{16,18,23}. However, prohormones have been found to be considerably less potent than their product²⁴. Thus, the present demonstration that SRIF-25 and SRIF-28 are about 10-fold more potent than SRIF-14 in inhibiting insulin *in vitro* and previous reports^{17,28} that synthetic replicates of these peptides are 10-fold more potent in inhibiting growth hormone secretion *in vitro* and *in vivo* suggest that SRIF-25 and SRIF-28 are not mere biosynthetic precursors for SRIF-14 in the classical sense of a prohormone but rather that they should be considered as alternate forms of biologically active somatostatins.

As even the secreted peptides may be further processed before interaction at receptors in target tissues²⁵, it will remain uncertain which peptide is the precursor and which the ultimate product until it is determined which peptide is actually bound at the SRIF receptor *in vivo*. Furthermore, it is possible that, as with proopiomelanocortin which is processed differently in adenophyseal cells and cells of the pars intermedia²⁶, prosomatostatin may also be processed differently in various tissues. In some tissues, SRIF-14 may represent the predominant secretory product whereas in others it may be SRIF-28 or SRIF-25. Moreover, the same cell may release different amounts of the peptides in different physiological or pathological conditions.

There is evidence that both SRIF-14 and SRIF-28 are present in the pancreas, presumably in islet D-cells²⁷. That SRIF-14 is a

Table 1 Effects of somatostatin-14, -28 and -25 on *in vitro* insulin and glucagon secretion

	Insulin (ng)	Glucagon (ng)
Control (12)	483 ± 13	57.7 ± 4.8
Somatostatin-14		
3.3 × 10 ⁻¹³ (14)	489 ± 28	39.9 ± 2.7*
3.3 × 10 ⁻¹² (12)	389 ± 26*	27.2 ± 1.3*
3.3 × 10 ⁻¹¹ (7)	308 ± 24*	19.8 ± 1.8*
1.7 × 10 ⁻¹⁰ (11)	227 ± 14*	12.2 ± 0.7*
Somatostatin-28		
3.3 × 10 ⁻¹⁴ (5)	449 ± 38	55.1 ± 5.7
3.3 × 10 ⁻¹² (8)	276 ± 20*†	40.0 ± 3.6*†
3.3 × 10 ⁻¹¹ (6)	210 ± 18*†	35.0 ± 1.1*†
1.7 × 10 ⁻¹⁰ (6)	157 ± 26*†	19.2 ± 2.0*†
Somatostatin-25		
3.3 × 10 ⁻¹⁴ (5)	452 ± 31	53.6 ± 1.5
3.3 × 10 ⁻¹² (6)	273 ± 22*†	40.8 ± 4.7*†
3.3 × 10 ⁻¹¹ (5)	239 ± 20*†	29.8 ± 2.1*†
1.7 × 10 ⁻¹⁰ (5)	166 ± 20*†	19.4 ± 2.9*†

Data represent the total amount of hormone released during each 15-min stimulation period (see text for details) and are given as mean ± s.e.m. Numbers in parentheses indicate the number of experiments.

* *P* < 0.05 against control.

† *P* < 0.05 against somatostatin-14.

* To whom reprint requests should be addressed.

more potent inhibitor of arginine-stimulated glucagon release than of insulin release has been previously noted²⁰. The present finding that SRIF-28 is a more potent inhibitor of arginine-stimulated insulin release than of glucagon release raises the possibility that preferential secretion of SRIF-14 and SRIF-28 may provide a novel mechanism for selective regulation of insulin and glucagon release. This suggestion is consistent with the observation that SRIF-28 seems to comprise less than 5% of SRIF-like immunoreactivity within the pancreas²⁷ because SRIF-28 is a 10-fold more potent inhibitor of insulin release than is SRIF-14, the predominant molecular species in the pancreas. It remains to be determined, however, whether the present observations regarding arginine-stimulated insulin and glucagon release are applicable to other secretagogues of these hormones.

This work was supported in part by grants AM20837, AM07147, AM07352 and AM11811 from the NIH and from the Mayo Foundation, Hearst Foundation and the National Foundation March of Dimes.

Received 2 February; accepted 9 March 1981.

1. Hokfelt, T. *et al. Acta endocr.* **80**, 1–41 (1978).
2. Parson, J., Erlandsen, S., Hegre, O., McEvoy, R. & Elde, R. *J. Histochem. Cytochem.* **24**, 872–882 (1976).
3. Yamada, Y., Ito, S., Matsubara, Y. & Kobayashi, S. *Tohoku J. exp. Med.* **122**, 87–92 (1977).
4. Renald, L., Martin, J. & Brazeau, P. *Nature* **255**, 233–235 (1975).
5. Cohn, M. & Cohn, M. *Brain Res.* **96**, 138–141 (1975).
6. Guillemin, R. & Gerich, J. A. *Rev. Med.* **27**, 379–388 (1976).
7. Creutzfeldt, W. & Arnold, R. *Metabolism* **27**, 1309–1316 (1978).
8. Raptis, S., Schlegel, W., Lehmann, E., Dollinger, H. & Zoupas, C. *Metabolism* **27**, 1321–1328 (1978).
9. Brazeau, P. *et al. Science* **179**, 77–79 (1973).
10. Schally, A. *et al. Biochemistry* **15**, 509–514 (1976).
11. Patel, Y. & Reichlin, S. *Endocrinology* **102**, 523–530 (1978).
12. Larsson, L. *et al. Lancet* **i**, 666–668 (1977).
13. Vale, W. *et al. Metabolism* **25**, 1491–1494 (1976).
14. McIntosh, C. *et al. Gut* **19**, 656–663 (1978).
15. Arimura, A., Sato, H., DuPont, A., Nishi, N. & Schally, A. *Science* **189**, 1007–1008 (1975).
16. Pradayrol, L., Jorvall, H., Mutt, V. & Ribet, A. *FEBS Lett.* **109**, 55–58 (1980).
17. Brazeau, P. *et al. C. R. heb. Séanc. Acad. Sci., Paris* **290**, 1369–1372 (1980).
18. Schally, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 4489–4493 (1980).
19. Ling, N. *et al. Biochem. biophys. Res. Commun.* **95**, 945–951 (1980).
20. Gerich, J., Lovinger, R. & Grodsky, G. *Endocrinology* **96**, 749–754 (1975).
21. Herbert, V., Lau, K., Gottlieb, C. & Bleicher, S. *J. clin. Endocr. Metab.* **25**, 1375–1384 (1965).
22. Faloona, G. & Unger, R. in *Methods of Hormone Radioimmunoassay* (eds Jaffe, G. & Behrman, H.) (Academic, New York, 1974).
23. Meyers, C., Murphy, W., Redding, T., Coy, D. & Schally A. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
24. Steiner, D. in *Peptide Hormones* (ed. Parsons, J.) (Macmillan, London, 1976).
25. Peart, W. *Pharmac. Rev.* **17**, 143–182 (1965).
26. Eipper, B. & Mains, R. *J. Supramolec. Struct.* **8**, 247–262 (1978).
27. Benoit, R., Bohlen, P., Ling, N. & Guillemin, R. *Endocrinology* **107**, 2127–2129 (1980).
28. Brazeau *et al. Reg. Peptides* **1**, 255–264 (1981).

γ Interferon induced by *S. aureus* protein A augments natural killing and ADCC

William J. Catalona, Timothy L. Ratliff & Robert E. McCool

The Yalem Research Laboratories of The Jewish Hospital of St Louis and the Division of Urology, Washington University School of Medicine, St Louis, Missouri 63110, U.S.A.

Staphylococcus aureus produces a protein identified as protein A (SpA) with a molecular weight of 41,000 (ref. 1) which binds to the Fc region of many types of mammalian IgG², activates complement³, blocks opsonins⁴ and is both chemotactic⁵ and mitogenic^{6,7}. SpA has been used for various immunological purposes including removal of immune complexes⁸, arming effector cells with antibody⁹, distinguishing between antibody-dependent cell-mediated cytotoxicity (ADCC) and natural killing (NK)¹⁰, and exerting anti-tumour activity¹¹. We have now demonstrated that SpA induces the production of γ interferon (IFN- γ) in nonadherent, T lymphocyte-depleted (E⁻), Fc receptor-bearing (FcR⁺) human peripheral blood lymphocytes. Our data also suggest that IFN- γ is a potent stimulator of both NK and ADCC activity.

Table 1 Antigenic and physicochemical characterization of SpA-induced IFN

	Treatment of supernatant	Interferon (unit ml ⁻¹)
Physicochemical characterization	None (mock interferon)	≤ 5
	None (interferon)	250
	Trypsin	≤ 5
	pH 2	≤ 5
Antigenic characterization	56 °C	≤ 5
	None	2,000
	Anti-hu IFN- α	2,000
	Anti-hu IFN- γ	100

Human PBLs were treated with SpA (50 μ g ml⁻¹) for 2 h, washed and incubated for an additional 24 h. Supernatants were collected as interferon preparations. Mock interferon was prepared in an identical fashion except that PBLs were treated with SpA diluent only. Supernatant treatments: trypsin: supernatants were treated with 100 μ g ml⁻¹ trypsin for 30 min at 37 °C; pH 2: supernatants were dialysed against glycine-HCl buffer (pH 2.0) for 24 h. The pH was returned to 7.2 by dialysis against minimal essential medium with Earl's salts; 56 °C: Supernatants were maintained at 56 °C for 1 h. Neutralization studies with antiserum against human (hu) IFN- α and hu IFN- γ were kindly performed by Dr John Stanton, University of Texas Medical Branch, Galveston, Texas.

Previous reports have shown that staphylococcal enterotoxin A (SEA), a T-cell mitogen, induces IFN- γ production¹². However, the interferon-producing potential of SpA has not previously been reported, and has not hitherto been a consideration in the interpretation of the biological properties of SpA.

In the present study, human peripheral blood lymphocytes (PBLs) were isolated by centrifugation of heparinized whole blood on Ficoll-Hypaque gradients as described previously¹³. Interface cells were washed three times with phosphate-buffered saline (PBS) and resuspended to the desired concentration in RPMI 1640 medium supplemented with 10% fetal calf serum and 1% antibiotic-antimycotic solution (Gibco) (complete medium). Interferon was induced by incubating PBLs (5 $\times$ 10⁶ ml⁻¹) with 50 μ g ml⁻¹ SpA (Pharmacia; no SEA contamination has been detected in SpA preparations from Pharmacia [James L. Richey (Pharmacia), personal communication]) for 2 h at 37 °C in a 5% CO₂ atmosphere. Following incubation with SpA, PBLs were washed three times with PBS, resuspended to 5 $\times$ 10⁶ cells ml⁻¹ in complete medium and incubated for 24 h at 37 °C. Supernatants were collected as interferon preparations and frozen at -70 °C until used. Mock interferon was prepared in an identical fashion except that SpA was not added to PBL cultures. Supernatants were assayed for interferon activity by measuring their capacity to inhibit the cytopathic effect of vesicular stomatitis virus on human foreskin fibroblast cultures (Biofluids). Titres are reported as International Reference units (NIH Human Leukocyte Reference Standard GO 23-902-527).

Interferon assays were performed on supernatants of SpA-treated PBLs from more than 20 donors. The interferon titres varied between donors, ranging from 100 to 300 U ml⁻¹ (mean = 160 $\pm$ 20 U ml⁻¹) using unfractionated PBLs. A representative experiment is shown in Table 1. The results show that supernatants from SpA-treated PBLs exhibited antiviral activity whereas none was present in mock interferon preparations. Physicochemical and antibody neutralization studies (Table 1) suggest that SpA-induced interferon is immune interferon (IFN- γ). Cell fractionation studies (Table 2) showed that depletion of adherent cells on glass surfaces and removal of the majority of T lymphocytes by E-rosette depletion enhanced interferon production, whereas depletion of FcR⁺ cells on antibody-sensitized sheep erythrocytes abrogated interferon production. These results suggest that with SpA, unlike other mitogen interferon inducers, macrophage depletion enhances

Table 2 Determination of the mononuclear cell subpopulation responsible for SpA-induced interferon production

Experiment no.	Cell fraction	Interferon (units ml ⁻¹)	% Latex positive	Cell markers % FcR positive	% E-rosette positive
1	Unfractionated	250	18	ND	ND
	Macrophage enriched	125	64	ND	ND
	Macrophage depleted	1,580	2	ND	ND
2	Macrophage depleted	625	ND	5	ND
	FcR negative	25	ND	1	ND
3	Macrophage depleted	200	ND	ND	70
	T-lymphocyte depleted	1,250	ND	ND	16

The measurement of latex ingestion, Fc receptors and E-rosette formation was performed as previously described¹³. PBLs were incubated on glass Petri plates for 1 h at 37 °C. Macrophage-depleted cells were removed by vigorous washing. Macrophage-enriched cells were removed from the glass with a rubber policeman. For FcR-negative cells, FcR-positive lymphocytes were removed by IgG-coated sheep erythrocytes as previously described¹³. For T lymphocyte-depleted fraction, PBLs were incubated with sheep erythrocytes overnight at 4 °C. After incubation T lymphocytes were removed by Ficoll-Hypaque density centrifugation as previously described¹³. ND, no data.

rather than reduces interferon production¹⁴ and further, that nonadherent, non-T, FcR⁺ lymphocytes produce interferon after SpA stimulation.

Interferons have been classified as IFN- α , IFN- β and IFN- γ , based on antigenic and physicochemical properties. IFN- α , - β and - γ are antigenically distinct compounds¹⁴. Furthermore, IFN- α and IFN- β , unlike IFN- γ , are stable to pH 2 and to heating at 56 °C for 1 h (ref. 15). All interferon types are inactivated by trypsin treatment. Our characterization studies revealed that SpA-induced interferon was inactivated by pH 2 treatment for 24 h or heating at 56 °C for 1 h. Moreover, treatment of SpA-induced interferon with antiserum to human IFN- α failed to neutralize its antiviral activity whereas antiserum to human IFN- γ succeeded, suggesting that SpA-induced interferon is IFN- γ .

Previous reports have shown that IFN- α and IFN- α inducers augment both ADCC and NK activity¹⁶. We therefore evaluated the effect of SpA-induced IFN- γ -containing supernatants on ADCC and NK activity (Fig. 1). Effector PBLs were treated with SpA-induced interferon-containing supernatants overnight at 37 °C (controls were treated with mock interferon) before the addition of ⁵¹Cr-labelled target cells. The cytotoxic activity of both NK directed against K562 human myeloid leukaemia target cells and ADCC directed against Chang human liver cell targets was augmented by interferon-containing supernatants (Fig. 1). Significant augmentation was observed

at effector/target cell ratios ranging from 6:1 to 50:1 (data not shown). Furthermore, trypsin digestion, heating at 56 °C and pH 2 treatment which inactivated the antiviral activity of SpA-induced interferon also significantly diminished that augmentation of cytotoxic activity (data not shown).

SpA has often been used in short-term cytotoxicity assays to distinguish between ADCC activity, which is inhibited by SpA, and NK activity which is not¹⁷. Kinetic studies on the rate of interferon production after SpA treatment showed that interferon activity was not apparent until 8 h after SpA treatment and peaked by 24 h (data not shown). Thus, only long-term cytotoxicity assays would be expected to reflect the interferon-induced augmentation of NK and ADCC activity. Indeed, Pape *et al.*¹⁸ recently reported that SpA-treated PBLs exhibited augmented NK activity against T 24 bladder cancer cells in an 18-h assay, but not against K562 cells in a 4-h assay. Other investigators have observed that SpA-treated effector cells exhibited augmented NK activity in 18-h, but not in short-term¹⁰, assays¹⁷.

Previous investigators have not considered the interferon-inducing potential of SpA in interpreting its diverse effects in biological systems. Cowan *et al.*⁸ added SpA to ADCC assays containing immune complexes and attributed the enhanced cytotoxicity observed to removal of the immune complexes. Sulica *et al.*⁹ found that effector cells treated with target cell-specific antibody complexes to SpA exhibited greater cytotoxic activity than effector cells treated with antibody alone. They attributed the enhanced cytotoxicity to increased antibody arming in the presence of SpA. Our results suggest that the enhancement of cytotoxicity observed by both Cowan *et al.*⁸ and Sulica *et al.*⁹ may be attributed, at least in part, to interferon induction.

Ray and co-workers¹¹ demonstrated that *S. aureus* (Cowan), a producer of SpA, has antitumour activity against a murine methylcholanthrene-induced fibrosarcoma. These investigators suggested that the antitumour activity was associated with augmentation of the humoral immune response to tumour antigens and removal of immune complexes. Again, our results suggest that SpA-induced interferon production may be implicated.

Thus, our results demonstrate that SpA induces IFN- γ production by nonadherent, T lymphocyte-depleted, FcR⁺ human PBLs, and that IFN- γ augments both NK and ADCC activity. Our results suggest that the induction of IFN- γ by SpA exhibits several unique characteristics. Interferon induction by SpA seems to be less dependent on macrophages than induction of IFN- γ by other mitogens¹⁴ and is apparently produced by non-T lymphocytes although more complete cell fractionation studies must be performed to confirm this preliminary observation. Maximum production of IFN- γ by SpA occurs within 24 h whereas other mitogens require 3–5 days for maximum IFN- γ production¹⁴. Finally, our data suggest that investigators

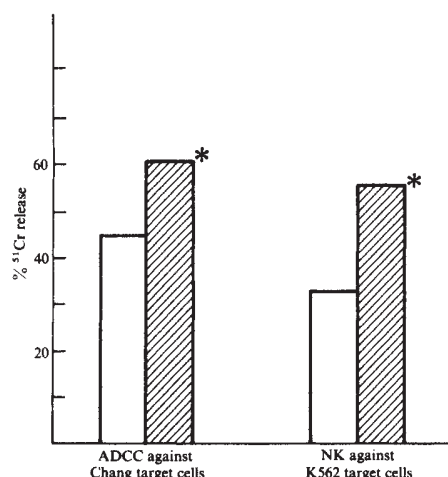


Fig. 1 Effect of SpA-induced interferon-containing supernatant on NK and ADCC activity of human PBLs. Both ADCC and NK assays were performed at 50:1 effector/target ratios as described previously¹³. Open bars represent untreated effector cells, hatched bars interferon-treated effector cells. *, $P < 0.005$ by Student's *t*-test.

using SpA as an immunological reagent for its Fc receptor-bearing properties must consider its interferon-producing potential in interpreting results.

We thank Joseph M. Davie, J. Russell Little Jr, Sondra Schlesinger and Gordon W. Philpott for critical reviews of this manuscript. We also thank Marsha Collett and Kimberly Merritt for help in preparing the manuscript. This work was supported in part by PHS grant 1 R26 CA23855, NCI grant 1 R26 CA28860 through the National Bladder Cancer Project, grant 1 R26 CA25792 from the National Prostate Cancer Project, Washington University Biomedical Research grant 3375-55874-B, General Research Support grant RIA-N50104-0 and Biomedical Research Support grant RIA-N90100-0 from The Jewish Hospital of St Louis, and American Cancer Society Institutional Research grant 1 N36Q. W.J.C. is recipient of American Cancer Society Junior Faculty Clinical Fellowship 382.

Received 8 April 1980; accepted 2 February 1981.

1. Bjork, I., Peterson, B.-A. & Sjoquist, J. *Eur. J. Biochem.* **29**, 579-584 (1972).
2. Kronvall, G. & Williams, R. C. *J. Immun.* **103**, 828-833 (1969).
3. Kronvall, G. & Gewurz, H. *Clin. exp. Immun.* **7**, 211-220 (1970).
4. Dossett, J. H., Kronvall, G., Williams, R. C. & Quie, P. G. *J. Immun.* **103**, 1405-1410 (1969).
5. Harvey, R. L., Kronvall, G., Troup, G. M., Anderson, R. F. & Williams, R. C. *Proc. Soc. exp. Biol. Med.* **135**, 453-456 (1970).
6. Forsgren, A., Svedjelund, A. & Wigzell, H. *Eur. J. Immun.* **6**, 207-213 (1976).
7. Sakane, T. & Green, I. *J. Immun.* **120**, 302-311 (1978).
8. Cowan, F. M., Klein, D. L., Armstrong, G. R. & Pearson, J. W. *Biomedicine* **30**, 23-27 (1979).
9. Sulica, A., Lakey, M., Gherman, M., Ghetie, V. & Sjoquist, J. *Scand. J. Immun.* **5**, 1191-1197 (1976).
10. Kay, H. D., Gonnard, G. D., West, W. H. & Herverman, R. B. *J. Immun.* **118**, 2058-2066 (1977).
11. Ray, P. K., Cooper, D. R., Bassett, J. G. & Mark, R. *Fedn Proc.* **38**, 1089 (1979).
12. Langford, M. P., Georgiades, J. A., Stanton, G. J., Dianzani, F. & Johnson, H. M. *Infect. Immun.* **26**, 36-41 (1979).
13. Catalana, W. J., Ratliff, T. L. & McCool, R. E. *J. Immun.* **122**, 1009-1014 (1979).
14. Epstein, L. B. *Tex. Rep. Biol. Med.* **35**, 42-56 (1977).
15. Wheelock, E. F. *Science* **169**, 310-311 (1965).
16. Herberman, R. B., Ortaldo, J. R. & Bonnard, G. D. *Nature* **277**, 221-223 (1979).
17. Vessella, R. L., Gormus, B. J., Lange, P. H. & Kaplan, M. E. *Int. J. Cancer* **21**, 594-603 (1978).
18. Pape, G. R., Troye, M., Axelsson, B. & Perlman, P. *J. Immun.* **122**, 2251-2260 (1979).

Production of immune interferon by murine T-cell clones from long-term cultures

Fabrizio Marcucci*, Michael Waller†, Holger Kirchner* & Peter Krammer†

* Institute of Virus Research and † Institute of Immunology and Genetics, German Cancer Research Centre, Heidelberg, FRG

Production of leukocyte interferon (IFN- α) and fibroblast interferon (IFN- β) can be induced by a variety of agents¹ but immune interferon, IFN- γ , is only obtained when lymphoid cells are stimulated by specific antigens, allo-antigens or T-cell mitogens^{2,3}. Moreover, in bulk cultures, only small quantities of IFN- γ are produced. The type of cell producing IFN- γ has not been unambiguously defined⁴⁻⁶ and so we set out to determine whether a pure T-cell population could produce it, exploiting the knowledge^{7,8} that T cells can be maintained indefinitely in tissue culture by the addition of T-cell growth factors. Although not all T cells can found long-term cultures of this kind⁹, cultures of antigen-specific helper, suppressor and killer T cells have been obtained¹⁰⁻¹⁵ in this way. We now describe the production of substantial amounts of IFN- γ when some (but not all) murine T-cell clones derived from such cultures are stimulated by either concanavalin A (Con A) or phytohaemagglutinin (PHA).

We obtained such a source of homogeneous T cells from clones of a 4-day culture of cells from the draining axillary and inguinal lymph nodes of AKR/J mice sensitized on the abdominal skin by 2,4,6-trinitrophenyl chloride (picryl chloride, PC) 5 days earlier¹⁴. These clones had been passaged in tissue culture for several months with a doubling time of ~20 h and grew strictly dependent on the presence of T-cell growth factors in the supernatant from rat spleen cells stimulated by Con A for 48 h. Cells from these clones were stimulated with Con A or PHA and tested for interferon production in different culture conditions.

The first series of experiments showed that these two mitogens induced interferon only when the T-cell growth factor-containing medium was replaced by fresh culture medium during interferon induction (Table 1). No antiviral activity could be detected without addition of T-cell mitogens or in the presence of B-cell mitogens (data not shown). The optimal dose for interferon induction, subsequently used for all further experiments was 5 or 10 $\mu\text{g ml}^{-1}$ Con A or PHA. Table 1 indicates that interferon was also generated by mitogen stimulation in serum-free medium and its production significantly increased by addition of 2-mercaptoethanol to the medium. The antiviral activity in the supernatant of the stimulated T-cell clones was characterized as IFN- γ by the following criteria (outlined in Table 2): (1) species-restricted activity, (2) induction of the antiviral state dependent on intact RNA synthesis in the target cell, (3) sensitivity to pH 2 and to trypsinization, (4) resistance to RNase treatment, (5) persistence of the antiviral activity after ultracentrifugation, (6) loss of activity after incubation with anti-IFN- γ antiserum and persistence of activity after incubation with anti-IFN- α and β antiserum.

Thus, production of IFN- γ could be induced by T-cell mitogens from continuously growing, nontransformed T cells maintained by passage in long-term culture in medium containing T-cell growth factors and in the absence of adherent cells. Table 3 shows that this was true for several randomly selected T-cell clones. Furthermore, while most clones responded equally well to stimulation with Con A or PHA, some produced more interferon on stimulation with one or the other. Few clones gave only marginal responses; most produced IFN- γ at titres considerably higher than those obtained in bulk cultures stimulated with the same mitogens (data not shown). The capacity to produce IFN- γ in response to mitogen stimulation seemed to be a stable property of the clones as approximately the same titres were obtained from two randomly selected clones tested after a 2-month interval (Table 3).

Our cultures had not been established by deliberate isolation and further growth of single cells but by seeding 1 cell per culture well. As 5.2% of the initial culture wells were positive for growth, most of the established long-term cultures from these wells could be considered clones according to Poisson statistics. The clones originated from a bulk culture of more than 95% T cells and grew strictly dependent on the presence of T-cell growth factor-containing medium. Furthermore, IFN- γ release could only be stimulated by T-cell mitogens.

To characterize the phenotype of one of the clones producing IFN- γ we treated the clone PC AKR 96, from the same series as the other clones, with the monoclonal anti-Thy 1.1, anti-Lyt 1.2 or anti-Lyt 2.1 antibodies and complement. As shown in Table 4, the cells were resistant to lysis, evaluated by Trypan blue exclusion, even after two consecutive treatments with any of the above antibodies. This insensitivity to complement treatment for T-cell clones in long-term culture has been described by others¹⁵. Antibody-treated and control cells were resuspended in T-cell growth factor-containing medium and after 4 days, growth (evaluated by ³H-thymidine uptake) and mitogen-induced IFN- γ production were determined. At this time, cells treated with anti-Thy 1.1 + complement incorporated only low levels of ³H-thymidine and almost no IFN- γ production could be observed. The phenomenon was specific because ³H-thymidine uptake and IFN- γ production were unaltered after treatment of the cells with the monoclonal anti-Thy 1.2 antibody,

Table 1 Conditions for IFN- γ production by T-cell clones stimulated with Con A or PHA

Clone	Mitogen added	$\mu\text{g ml}^{-1}$	IFN- γ titres in different culture conditions (International Units per ml)			
			NCM	TCGFCCM	SF-NCM	NCM + 2-ME
PC AKR 29	None		<3	30	<3 *	10
	PHA	5	240	45	300	550
	PHA	10	550	30	550	720
	Con A	5	240	30	240	550
	Con A	10	720	45	720	1,000
PC AKR 54	None		<3	30	<3	5
	PHA	5	350	30	500	720
	PHA	10	240	45	240	500
	Con A	5	550	45	550	720
	Con A	10	300	45	300	450

NCM, normal culture medium, RPMI-1640 (Gibco), supplemented with L-glutamine (2 mM), streptomycin ($100 \mu\text{g ml}^{-1}$) + 5% fetal calf serum (FCS). TCGFCCM, T-cell growth factor-containing conditioned medium, as NCM with 2-mercaptoethanol (2-ME, $3 \times 10^{-5}\text{M}$) and substituted with 10 mM HEPES, 10% FCS and 10% supernatant containing T-cell growth factors. This supernatant was obtained by stimulation of 5×10^6 Sprague-Dawley rat spleen cells (70 ml in Falcon bottles (flat), 3024 F) with $5 \mu\text{g}$ Con A for 24 h. The stimulation was done in medium with the same constituents as above except that Iscove's modified Dulbecco's medium without HEPES was used. SF-NCM, serum-free normal culture medium. T-cell clones were obtained in the following way. The draining axillary and inguinal lymph node cells of AKR/J mice sensitized 5 days previously on the abdominal skin by PC were put into culture as described in ref. 14. On day 5 of culture, cells cleared from dead cells and debris by centrifugation over Ficoll-Urovison¹⁴ were cloned by seeding 1 cell per well in $200 \mu\text{l}$ TCGFCCM containing 3×10^4 3,300 rad-irradiated AKR/J peritoneal exudate filler cells in 96-well flat-bottom microtitre plates (Linbro no. 76-003-05, Flow Laboratories). The old TCGFCCM was replaced once a week. After about 4 weeks all growing clones were separately transferred into 24-well plates (Linbro no. 76-033-05, Flow Laboratories) and after an additional 2 weeks the surviving clones were passaged and usually subcultured once a week in Falcon bottles (3024 F, Falcon) in TCGFCCM. All cultures were kept at 37°C , in humidified atmosphere, 95%–5% air– CO_2 mixture. For interferon induction, cells grown in TCGFCCM were centrifuged, washed twice and resuspended in NCM or in SF-NCM, or in NCM+2-ME at a concentration of $1 \times 10^6 \text{ ml}^{-1}$ and stimulated with PHA or Con A. Culture supernatants were collected 24 h later and interferon was titrated by means of a plaque reduction assay using L cells and vesicular stomatitis virus as challenge virus¹⁸.

which recognizes the allelic specificity of the Thy-1 locus not expressed on T cells from AKR mice. Typing of the Lyt markers for which monoclonal reagents were available (Lyt 1.2 and Lyt 2.1) was negative in all the assays performed. Thus, cells from this clone either are Lyt 1⁺2⁺ or possess these surface antigens at low density and are therefore not detectable in our experimental conditions. Clones negative for the expression of these markers have already been described^{15,16}. Bearing in mind that cultivation of long-term T-cell clones in T-cell growth factor-containing medium is accompanied by change in quantity or loss of expression of cell-surface antigens, particularly of the Ly series¹⁵, these observations suggest that the cells from the clone PC AKR 96 are derived from a single T cell and continue to express the θ antigen.

The clones had been passaged in tissue culture for more than 3 months without addition of adherent cells before they were tested for IFN- γ production. We therefore assume that they were virtually free of macrophages and that IFN- γ was being produced by T-cell clones without help from adherent cells. That some clones yielded titres of IFN- γ considerably higher than those normally obtained from cell mixtures in bulk cultures stimulated with the same mitogens indicates that, in these conditions either we obtained an expansion of the IFN- γ -producer cells or some regulatory mechanism acting in bulk cultures was bypassed.

When cultured in the absence of growth factor-containing medium, T-cell clones produced IFN- γ after mitogen stimulation (Table 1). Unstimulated cells from these cultures did not

Table 2 Characterization of antiviral activity in the supernatant of murine T-cell clone PC AKR 29 stimulated with Con A or PHA

Treatment of supernatant	Activity* on mouse cells
pH 2 for 24 h	–
Centrifugation at 100,000g for 1 h	+
Trypsin ($100 \mu\text{g ml}^{-1}$) for 1 h at 37°C	–
RNase ($1000 \mu\text{g ml}^{-1}$) for 1 h at 37°C	+
Incubation with anti-mouse IFN- α and - β antiserum for 1 h at 37°C	+
Incubation with anti-mouse IFN- γ antiserum for 1 h at 37°C	–

Anti-mouse IFN- α and - β antiserum was a rabbit anti-L cell interferon antiserum with a neutralization titre of 1:6,000 on IFN- α and - β . Anti-mouse IFN- γ antiserum was a rabbit antiserum raised against PHA-induced, mouse spleen cell interferon with a neutralization titre of 1:640, ref. 19. +, Activity undiminished; –, $>90\%$ loss of activity.

* Loss of antiviral activity was observed after pretreatment of L cells with $1 \mu\text{g ml}^{-1}$ actinomycin D for 1 h. The antiviral activity was species restricted as no reduction of virus plaques was observed on Rita (monkey kidney) cells.

Table 3 IFN- γ production by different T-cell clones stimulated with Con A or PHA

T-cell clones	IFN- γ titres (IU ml^{-1}) after stimulation with			
	PHA		Con A	
	$5 \mu\text{g ml}^{-1}$	$10 \mu\text{g ml}^{-1}$	$5 \mu\text{g ml}^{-1}$	$10 \mu\text{g ml}^{-1}$
PC AKR 28	500	300	60	80
PC AKR 37	15	10	15	30
PC AKR 41	300	240	240	240
PC AKR 74	150	100	500	800
PC AKR 75	50	15	240	100
PC AKR 81	90	30	150	100
PC AKR 83	720	600	800	720
PC AKR 89	30	15	15	30
PC AKR 96	250	400	60	100
PC AKR 45/1*	240	800	30	80
PC AKR 45/2*	300	720	20	60
PC AKR 59/1*	720	800	720	720
PC AKR 59/2*	500	600	600	720

Stimulation of T-cell clones was done in NCM (see Table 1 legend).

* 1 and 2 represent the results from clones PC AKR 45 and 59 at a time interval of 2 months.

Table 4 Analysis of cell-surface markers of the T-cell clone PC AKR 96

Treatment	Characterization of the cells after treatment			
	Day 0	Day 4		
	% Living cells	IFN- γ production (IU ml ⁻¹)	Growth (³ H-TdR incorporation, c.p.m.)	IFN- γ production (IU ml ⁻¹)
Control	90	1,000	148,520	600
Complement (C') alone	80	800	139,343	600
Anti-Thy 1.1+C'	75	720	3,047	10
Anti-Thy 1.2+C'	70	720	122,250	720
Anti-Lyt 1.2+C'	80	1,000	121,342	800
Anti-Lyt 2.1+C'	70	800	156,528	600

The control was NCM alone. Antibodies were monoclonal antibodies anti-Thy 1.1, anti-Lyt 1.2, Lyt 2.1 (code NEJ-014, NEJ-017 and NEJ-004, respectively, NEN) and anti-Thy 1.2 (code FfB5, Olac), and were used at a final dilution of 1:100, 1:50, 1:50 and 1:100, respectively. Complement was 1:6 diluted, selected rabbit complement. Treatment with antibodies and complement was performed twice with 2×10^6 cells as previously described²⁰. At the end of the treatment the cells were resuspended in 1 ml NCM. 0.2 ml of each cell suspension was induced for IFN- γ production with PHA (10 μ g ml⁻¹). 0.2 ml of a 1:10 dilution of each cell suspension were separately transferred into 24-well plates (Linbro no. 76-033-05, Flow Labs), each well containing 1 ml fresh TCGFCCM. 4 days later 3×0.2 ml of each cell suspension were separately transferred into a 96-well plate, pulsed for 4 h with 1 μ Ci ml⁻¹ of tritiated thymidine (³H-TdR, Amersham) after which the cultures were aspirated on filter paper by use of a semi-automated collecting device (Skatron, Flow Labs) and radioactivity of the filters measured in a scintillation counter (Mark II, Searle-Nuclear, Chicago). Another 0.2-ml aliquot was centrifuged and resuspended in 0.2 ml NCM containing PHA (10 μ g ml⁻¹). 24 h later the supernatants were tested for IFN- γ as described in Table 1 legend. Per cent living cells was determined by Trypan blue exclusion. Growth results are the mean value of triplicate samples.

release antiviral activity. Supernatants from unstimulated cells cultured with T-cell growth factor-containing medium were positive for IFN- γ because the medium itself contained IFN- γ . Addition of mitogens to cells from these cultures did not increase IFN- γ titres above the medium background (Table 1). This might have been due to blocking activity exerted by IFN- γ similar to the blocking observed for IFN- α and β production¹⁷.

T-cell help, cytotoxic activity and suppression can be obtained from T-cell lines and clones without antigen or mitogen stimulation¹⁰⁻¹². Our data indicate that other T-cell activities, like IFN- γ production, can be stimulated from T-cell lines and clones only with mitogens. It remains to be determined whether other stimuli, such as specific antigens, which resemble more closely the *in vivo* situation, are also able to induce IFN- γ production from T-cell clones.

Finally, the fact that large quantities of IFN- γ can be produced in serum-free conditions (Table 1) from homogeneous cells in continuous culture could be useful for its biochemical isolation.

We thank Professor K. Resch and Drs S. Landolfo and M. Simon for helpful discussions, Ms A. Michnay and M. Nowak for technical assistance and Ms M. Kasamasch for typing the manuscript. Drs June Dunnick and Juana Wietzerbin provided the anti-mouse IFN- α and - β antiserum and anti-mouse IFN- γ antiserum respectively.

Received 13 October 1980; accepted 25 February 1981.

- Stewart, W. E. II in *The Interferon System*, 25-27 (Springer, New York, 1979).
- Gifford, G. E., Tibor, A. & Peavy, D. L. *Infect. Immun.* **3**, 164-166 (1971).
- Wheelock, E. F. *Science* **149**, 310-311 (1965).
- Sonnenfeld, G., Mandel, A. D. & Merigan, T. C. *Immunology* **36**, 883-890 (1979).
- Neumann, C. & Sorg, C. *Eur. J. Immun.* **8**, 582-589 (1978).
- Kirchner, H. et al. *Eur. J. Immun.* **9**, 824-826 (1979).
- Gillis, S. & Smith, K. A. *Nature* **268**, 154-156 (1977).
- Morgan, D. A., Ruscetti, F. W. & Gallo, R. C. *Science* **193**, 1007-1008 (1976).
- V. Boehmer, H., Haas, W., Pohlitz, H., Hengartner, H. & Nabholz, M. *Springer Semin. Immunopath.* **3/1**, 23-27 (1980).
- Watson, J. D. J. *Exp. Med.* **150**, 1510-1519 (1979).
- McVay-Boudreau, L., Fresno, M., Nabel, G. & Cantor, H. *Abstr. 4th int. Congr. Immun.*, Paris, 4.2.14 (1980).
- Baker, P. E., Gillis, J. & Smith, K. A. *J. exp. Med.* **149**, 273-278 (1979).
- Nabholz, M., Engers, H. D., Collavo, D. & North, M. *Curr. Topics Microbiol.* **81**, 176-187 (1978).
- Krammer, P. H., Rehberger, R. & Eichmann, K. J. *exp. Med.* **151**, 1166-1182 (1980).
- Nabholz, M. et al. *Immun. Rev.* **51**, 125-156 (1980).
- Glasebrook, A. L. & Fitch, F. W. J. *exp. Med.* **151**, 876-895 (1980).
- Stewart, W. E. II, Grosser, L. B. & Lockart, R. Z. Jr *J. gen. Virol.* **13**, 35-50 (1971).
- Campbell, J. B., Gruberger, T., Kochmann, M. A. & White, S. L. *Can. J. Microbiol.* **3**, 1247-1253 (1975).
- Stefanos, S., Catinot, L., Wietzerbin, J. & Falcoff, E. J. *gen. Virol.* **50**, 225-229 (1980).
- Simon, M. M. & Abenhardt, B. *Eur. J. Immun.* **10**, 334-341 (1980).

Herpesvirus-transformed cytotoxic T-cell lines

Donald R. Johnson & Mikael Jondal

Department of Tumor Biology, Karolinska Institutet, S-104 01 Stockholm, Sweden

Investigations of cellular cytotoxicity of the immune system are hampered by the lack of continuously growing, transformed cell lines which express a cytotoxic potential. Here we describe cytotoxic cell lines from the cotton-topped marmoset monkey, transformed by Herpesvirus Ateles (HVA) or Herpesvirus Saimiri (HVS), which can kill certain target cells in a short-term *in vitro* test. HVA/HVS-transformed cells have earlier been classified as belonging to the T-cell lineage¹⁻³ in contrast to Epstein-Barr virus (EBV)-transformed cells derived from B lymphocytes⁴. We suggest that the HVA/HVS-transformed killer cell lines described here represent an effector population resembling, or corresponding to, marmoset natural killer (NK) cells and that they may be used to define the cytolytic mechanism involved in cellular cytotoxicity and possibly also effector cell receptors and target cell antigens, as well as regulatory mechanisms of general biological interest.

Peripheral blood lymphocytes from adult cotton-topped marmosets were separated on Ficoll-Isopaque and transformed by cell-free supernatants from HVA-infected marmoset kidney fibroblasts³. The resulting cell lines were propagated as suspension cultures in RPMI 1640 media supplemented with 10% fetal calf serum and antibiotics. A 6-h ⁵¹Cr release cytotoxicity assay was performed in microplates as described in Table 1 legend, using various target cells. Cytotoxicity assays in which HVA-carrying cell lines were used as effector cells, were compared with cytotoxicity expressed by marmoset and human NK cells isolated from the peripheral blood of normal healthy donors. Five different HVA-transformed cell lines and two HVS-transformed cell lines were tested for cytotoxic activity against the human T-cell line Molt-4 (Table 1) derived from acute lymphocytic leukaemia⁵. All seven HVA/HVS-transformed cell lines, and human NK cells, were cytotoxic in contrast to human fetal thymocytes and marmoset B-cell lines. Both *in vitro*-transformed cell lines and lines transformed *in vivo*, that is, tumour-derived cells, were cytotoxic, which indicates that the cytotoxic

Table 1 Cytotoxicity by HVA/HVS-transformed cell lines from cotton-topped marmoset monkeys

Effector cell	Transforming virus	Origin	Type	Age	Cytotoxicity against target cell Molt-4 (specific ⁵¹ Cr release)		
					25:1	12.5:1	6.3:1
Experiment I							
77-DI-2	HVA	<i>In vitro</i> transformed	T	6 months	68.4	59.9	48.9
70-KI-1	HVA	Same	T	3 yr	60.1	49.6	38.2
22-CM-37	HVA	Same	T	4 yr	11.4	9.8	7.9
22-CM-61	HVA	Same	T	4 yr	32.2	29.6	19.8
1022	HVA	Tumour-derived	T	6 yr	62.3	52.1	44.8
70-N-2	HVS	Same	T	8 yr	52.1	40.8	33.8
MLC	HVS	Same	T	9 yr	19.2	15.3	10.7
Human NK	—	Peripheral blood	—	Fresh	68.7	48.3	28.4
Human thymocytes	—	Fetal thymus	—	Fresh	0.2	-0.1	0.0
Experiment II							
77-DI-2	HVA	<i>In vitro</i> transformed	T		66.2	48.7	39.4
77-DI-2	EBV	Same	B		3.1	0.6	0.1
70-KI-1	EBV	Same	B		0.2	0.1	0.2
Human NK	—	Peripheral blood	—		75.4	63.6	54.1

Cytotoxicity assay was performed in V-shaped microtitre plates using 10^4 ⁵¹Cr-labelled target cells. RPMI 1640 with 10% fetal calf serum was used as testing medium, giving a total volume of 150 μ l per well. After 6 h, 50 μ l was collected from each well and cytotoxicity calculated as follows: test release - spontaneous release/80% of maximum label - spontaneous release $\times$ 100. Values are means of duplicate wells; two representative experiments are shown.

potential of these cell lines is a stable characteristic that does not change even after many years in tissue culture (note particularly the old cell lines 1022, MLC and 70-N-2).

As the Molt-4 cell is a standard target for human NK cells⁶, and as the HVA/HVS cells are spontaneously cytotoxic (that is, no special triggering is needed), we investigated whether they would have the same target specificity as human NK cells when tested against an extended panel of cells (see Table 2). HVA/HVS cell lines were found to kill most efficiently target cells derived from T-cell malignancies, such targets are also readily killed by human NK cells⁷. The marmoset lines killed K562 but to a lesser degree than human NK cells (K562 is also a standard target for the human NK system^{6,8}). HVA/HVS cell lines killed B-cell lines from non-EBV (as opposed to EBV-containing) malignancies. In contrast, human NK cells seem to kill malignant as opposed to non-malignant cells, as previously shown⁹. Neither HVA/HVS lines nor human NK cells kill normal mitogen-transformed T cells—this was also confirmed using T blasts from concanavalin A and MLC cultures (data not shown). Finally, the HVA/HVS cell lines did not kill the standard NK target in the mouse system, YAC-1, a Moloney leukaemogenic virus (MLV)-transformed cell line of thymic origin^{8,10}.

Comparison of NK cells from cotton-topped marmosets with human NK cells (Table 3) showed that marmoset NK cytotoxicity resembled that of the 77-DI-2 cell line as both predominantly killed the Molt-4 target and, to a much lesser degree, the Raji and K-562 targets. In contrast, human NK cells demonstrated highest activity against the K-562 target. These results indicate that the difference in target cell specificity between the 77-DI-2 cell line and human NK cells (Table 2) reflects minor differences between the human and marmoset NK systems.

To ascertain that the cell line-mediated cytotoxicity is a true membrane-mediated event (as are conventional T- and NK-cell killing) and not dependent on secreted lymphotoxin-like molecules, we investigated cytotoxicity on the single cell level using a recently described technique in which separate effector-target cell conjugates were embedded in agarose¹¹. We also compared the number of active killer cells in the 77-DI-2 cell line with those present in normal marmoset lymphocytes propagated by T-cell growth factor (TCGF). TCGF was prepared from supernatants of phytohaemagglutinin (PHA)-stimulated human lymphocytes, usually pooled from three different donors, as described elsewhere¹². Table 4 shows that 38.5% of the HVA-transformed killer cells were able to bind to the target cells but

Table 2 Target cell specificity of the 77-DI-2 effector cell line

Target cell	Origin	Characteristics	EBV	Effector cells (specific ⁵¹ Cr release)	
				77-DI-2	Human NK
Molt-4	Acute lymphocytic leukaemia	T cell	—	68.9	68.7
1301	Same	Same	—	49.3	66.4
CCRF-CEM	Same	Same	—	71.2	68.9
HSB-2	Same	Same	—	71.6	79.7
HD-Mar	Hodgkin's disease	Same	—	36.5	94.8
BJAB	Burkitt's lymphoma	B cell	—	29.1	94.0
Ramos	Same	Same	—	15.9	20.6
K562	Chronic myelocytic leukaemia	CML cell	—	17.3	87.5
DHL-9	Diffuse histiocytic lymphoma	B' cell	—	29.4	52.8
Daudi	Burkitt's lymphoma	Same	+	-0.3	42.8
P3HR-1	Same	Same	+	0.1	65.4
B95-8	Normal marmoset B cells	Same	+	6.3	41.8
LCL(IE)	Normal human peripheral blood	Same	+	0.4	1.2
Human T-cell blasts	Normal human T cells stimulated by PHA	T cell	—	0.2	1.6
YAC-1	Mouse line, MLV-transformed	Same	—	0.7	9.9

Cytotoxicity assay was performed as described in Table 1 legend. The data represent three different experiments performed in identical conditions using an effector/target ratio of 25:1.

only 18.3% of these were active killer cells. These values were comparable with those found for the two TCGF-maintained cultures, which demonstrates equivalent cytotoxic potential for these two effector cell types.

TCGF has been shown to sustain proliferation of antigen-specific T killer¹³ cells but also to generate NK-like activity¹⁴, and it has been suggested that specific killing and NK killing may be expressed simultaneously in activated T cells at certain stages in T-cell differentiation¹⁵. Thus, because of uncertainties about the true nature of TCGF cultures and their instability, it seems that HVA/HVS-transformed cytotoxic cell lines are useful, as viral transformation probably freezes effector cells in a defined stage of differentiation representative of the cytotoxic potential expressed by the corresponding non-transformed cells. An analogy could be drawn here between HVA/HVS and EBV because EBV only transforms B lymphocytes in a defined stage of their differentiation⁴.

We found marked variation in the number of active killer cells in the 77-DI-2 cell line when tested at different time points (data not shown), the results in Table 4 being of average activity. The explanation for these variations is unclear, suggesting that regulatory mechanisms occur during tissue culture, but we know only that fast growing cultures maintained in roller bottles show the highest activity. The possibility of artificial killing by mycoplasma was excluded by comparing the activity of 77-DI-2 cells kept in mycoplasma-free conditions with that of cells in ordinary

Table 3 Natural cytotoxicity by peripheral blood lymphocytes from cotton-topped marmoset monkeys

Effector cells	Cytotoxicity against target cell (specific ⁵¹ Cr release)		
	1301	K-562	Raji
Monkey 1	43.4	12.1	11.2
Monkey 2	53.8	9.2	10.6
Monkey 3	33.9	8.2	6.2
77-DI-2 cell line	56.2	5.6	4.1
Human NK cells	31.5	72.3	9.3

Cotton-topped marmoset peripheral blood was obtained from Dr L. Wolfe of the Rush Medical Center, Chicago, USA. Lymphocytes were isolated on arrival and tissue-cultured overnight before testing. The cytotoxicity assay was performed using an effector/target cell ratio of 10:1 as described in Table 1 legend. Similar results were obtained in three separate repeat experiments.

tissue culture—similar cytotoxic activities were recorded (data not shown).

Morphological analysis of the 77-DI-2 cell line revealed that a substantial part of the cells resembled large granular lymphocytes (LGLs) and that this fraction seemed to be conjugated to susceptible target cells (data not shown). This may be significant as human NK cells have been reported as having LGL morphology⁸. Electron microscopy of 77-DI-2 cells showed immature cells with abundant endoplasmic reticulum and uropod formation. This indicates a secretory function and a high degree of mobility (data not shown).

Results from electron microscopy and the high cytotoxic activity on some target cells in ⁵¹Cr release assays compared with the number of active killer cells found in the agarose single cell cytotoxicity assay, led us to ask the question whether 77-DI-2 effector cells were able to recycle, that is, to kill more than one target cell in the given test period. We previously found this to be a characteristic of interferon-activated human NK cells^{16,17}. Using the same technique as for the latter to demonstrate recycling, we found a marked recycling capacity in the 77-DI-2 cell line (data not shown). This recycling capacity is also characteristic of immune cytotoxic T cells¹⁸.

In conclusion, we have found that HVA/HVS-transformed T-cell lines from cotton-topped marmosets can kill predominantly human T-cell lines and non-EBV B-cell lines of

Table 4 Cytotoxicity by the 77-DI-2 effector cell line and by TCGF-propagated normal marmoset lymphocytes

	Cytotoxicity (specific ⁵¹ Cr release)			TBCs* (%)	Cytotoxic TBCs† (%)
	3.0:1	1.5:1	0.8:1		
77-DI-2 (HVA)	50.2	40.7	31.6	38.5	18.3
77-DI-2 (TCGF)	58.6	52.5	50.7	28.7	23.1
77-EC-1 (TCGF)	72.9	65.6	56.2	31.8	30.4

Cytotoxicity was performed as described in Table 1 legend. Molt-4 was used as the target cell.

* Target binding cells (TBCs) were assayed as described elsewhere^{17,18}.

† Cytotoxic TBCs were tested by the agarose single cell cytotoxicity assay¹¹. Briefly, effector and target cells were conjugated at a ratio of 1:10 using carboxyfluorescein-diacetate-labelled target cells, and layered in agarose on to plastic Petri dishes. TBCs were estimated at 0 h and cytotoxicity at 6 h.

malignant origin. This reactivity is clearly a cell-mediated event requiring cell contact and not caused by lymphotoxin-like molecules. The specificity of HVA/HVS cell line killing resembles that of the marmoset NK system, and may be due to these viruses transforming marmoset NK cells or a subpopulation of them. If so, the low reactivity against EBV-induced tumours could indicate that marmoset NK cells differ from human NK cells in their target cell specificity. The selective EBV target deficiency of the marmoset NK system could possibly explain the high degree of susceptibility to EBV-induced lymphoproliferation/lymphomas in these animals⁴. We are now investigating this question.

Alternatively, the HVA/HVS cell line killing could be unrelated to the NK system and represent either a hitherto undefined effector population or possibly a T-effector population which is continuously generated in an autoreactive MLC type of culture. In either case, we consider that the HVA/HVS-transformed cytotoxic cell lines described here are the first immortalized cell lines of lymphoid origin reported which have the capacity to kill by cell contact, without added cell agglutinants, and which display a defined specificity. Possibly they represent an immune function of biological significance and will be useful for studying the finer mechanisms of cell-mediated cytotoxicity reactions.

We thank Maj-Britt Alter for technical assistance, Dr L. Wolfe for a continuous supply of marmoset blood and Dr L. Falk for supplying several of the investigated HVA/HVS cell lines. We also thank Dr G. Klein for discussing the data and manuscript. This work was supported by funds from the Swedish Cancer Society, Karolinska Institute and Karolinska Hospital and by contract NO1-CP-33316 from the Division of Cancer Cause and Prevention, NCI. D. R. J. is a recipient of a fellowship from the Cancer Research Institute, New York.

Received 25 November 1980; accepted 9 March 1981.

- Deinhardt, F. & Deinhardt, J. in *The Epstein-Barr Virus* (eds Epstein, M. A. & Achong, B. G.) 373-415 (Springer, New York, 1979).
- Falk, L., Wright, J., Wolfe, L. & Deinhardt, F. *Int. J. Cancer* **14**, 244-251 (1974).
- Falk, L., Johnson, D. R. & Deinhardt, F. *Int. J. Cancer* **21**, 652-657 (1978).
- Miller, G. in *The Epstein-Barr Virus* (eds Epstein, M. A. & Achong, B. G.) 351-372 (Springer, New York, 1979).
- Minowada, J., Ohnuma, T. & Moore, G. E. *J. natn. Cancer Inst.* **49**, 891-895 (1972).
- Jondal, M. & Pross, H. *Int. J. Cancer* **15**, 596-605 (1975).
- Ono, A., Amos, D. B. & Koren, H. S. *Nature* **266**, 546-548 (1977).
- Möller, G. (ed.) *Immun. Rev.* **44** (1979).
- Jondal, M., Spina, C. & Targan, S. *Nature* **272**, 62-64 (1978).
- Kiessling, R., Klein, E. & Wigzell, H. *Eur. J. Immun.* **5**, 112-117 (1975).
- Grimm, E. A., Thoma, J. A. & Bonavida, B. *J. Immun.* **123**, 2870-2877 (1979).
- Bonnard, G. D., Ysaka, K. & Maca, R. D. *Cell. Immun.* **51**, 390-401 (1980).
- Möller, G. (ed.) *Immun. Rev.* **51** (1980).
- Massucci, M. G., Klein, E. & Argov, S. *J. Immun.* **124**, 2458-2463 (1980).
- Klein, E. *Immun. Today* **1**, 4-6 (1980).
- Ullberg, M. & Jondal, M. *J. exp. Med.* **153**, 615-628 (1981).
- Ullberg, M., Merrill, J. & Jondal, M. *Scand. J. Immun.* (in the press).
- Berke, G. *Prog. Allergy* **27**, 69-116 (1979).

N-acetoxy-acetylaminofluorene reacts preferentially with a control region of intracellular SV40 chromosome

Peter Beard*, Motohisa Kaneko† & Peter Cerutti‡

Departments of *Virology and †Carcinogenesis, Swiss Institute for Experimental Cancer Research, 1066 Epalinges, Switzerland

Many chemical carcinogens or their metabolites react with DNA^{1,2}; thus it is of interest to determine what effect chromosomal structure has on these reactions. The chromosome of simian virus 40 (SV40) is well suited for such studies; like chromatin of eukaryotic cells, it is organized into nucleosomes. The nucleotide sequence of SV40 is known^{3,4}, together with much about the pattern of viral gene expression and DNA replication⁵, and the structure of the viral chromosome⁶⁻¹⁴. We have investigated the binding of the ultimate carcinogen, N-acetoxy-acetylaminofluorene (AAAF), to specific regions of the SV40 chromosome *in situ* in the intact infected cell. The results, reported here, indicate that a region containing regulatory functions on the intracellular SV40 chromosome has unique structural properties which render it more susceptible to attack by AAAF than the rest of the SV40 genome. The preferential binding of AAAF to regulatory regions of chromatin may have implications for the mechanism of action of this and similar carcinogens.

AAAF binds covalently to DNA both *in vitro* and in cells growing in culture¹⁵⁻¹⁸. Over 95% of the reaction products involve guanine residues^{15,16,18}. Like other carcinogens which introduce bulky substituents to DNA bases, AAAF reacts more extensively with internucleosomal linker DNA than with nucleosomal core DNA¹⁸⁻²².

We infected monolayer cultures of the monkey kidney cell line BSC-B with SV40. In the late phase of the lytic cycle, during viral DNA replication, the medium was replaced with medium containing ³H-AAAF (see Table 1 for details). After labelling for 15 min at 37 °C the cells were washed and SV40 DNA selectively extracted²³ and purified by centrifugation in gradients of CsCl-ethidium bromide²⁴. As a control we treated protein-free SV40 DNA *in vitro* with ³H-AAAF in the following way. SV40 DNA was extracted from infected cells by the method described above. The supercoiled DNA was then treated with DNA topoisomerase²⁵ to relax superhelical turns.

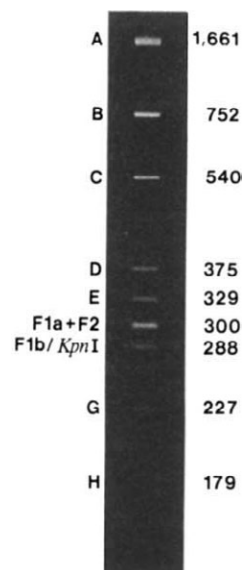


Fig. 1 Low-gelling-temperature agarose gel electrophoresis of SV40 DNA after digestion with restriction enzymes *HaeIII* plus *KpnI*. DNA was electrophoresed in a 3% gel of low-gelling-temperature agarose (Sigma) in Tris-phosphate-EDTA buffer, pH 7.8, for 6 h at 4 V cm⁻¹. The gel was stained with ethidium bromide (1 µg ml⁻¹) and photographed in light of 254 nm wavelength. Fragments resulting from cleavage by *HaeIII* alone are named according to Lebowitz *et al.*³⁸. The larger fragment produced by *KpnI* cleavage of *HaeIII* fragment F1b is marked *HaeIII* F1b/*KpnI*; the smaller product of this cleavage is too short to be detected in this gel (see Fig. 2c). Fragment lengths in base pairs are from ref. 5.

Table 1 Distribution of radioactivity in SV40 DNA restriction fragments after labelling of DNA *in vitro* or intracellularly with ³H-AAAF

SV40 <i>HaeIII</i> fragments	Per cent of total guanines in each fragment	Per cent of total radioactivity in each fragment	
		Labelled as free DNA	Labelled <i>in situ</i>
A	30	29	27
B	15.9	18.1	16.1
C	11.1	11.6	9.7
D	7.4	6.5	8.6
E	7.4	6.5	6.6
F1a+F2	12.0	12.3	12.7
F1b/ <i>KpnI</i>	7.2	7.6	11.5
G	4.9	4.6	4.4
H	3.9	3.5	3.3

Intracellular SV40 DNA was labelled by addition of ³H-AAAF (9 µM, 878 mCi mmol⁻¹, Midwest Research Institute, Kansas City) to virus-infected cell cultures during 15 min at 40 h after infection. SV40 DNA was then extracted and purified^{23,24}. Free DNA (100 µg ml⁻¹) was labelled in solution in 10 mM Tris-HCl, 1 mM EDTA pH 7.5. Unincorporated AAAF was then removed by repeated extractions with ether and precipitation of the DNA with ethanol. Labelled DNA (2–10 × 10³ c.p.m.) was digested with endonucleases *HaeIII* plus *KpnI* (New England Biolabs) and electrophoresed in preparative gel in conditions described in Fig. 1 legend. To determine the radioactivity in each restriction fragment, a simple and reliable procedure was developed. Slices of low-gelling-temperature agarose containing the DNA bands were cut, and water added to 1 ml. The agarose was melted at 70 °C for 5 min and 15 ml Aquasol (NEN) were added and the suspension mixed briefly until homogeneous. Samples were cooled in the dark for several hours, then counted in a scintillation spectrometer. Gel slices with no DNA gave the same background values as Aquasol alone (30 c.p.m.). The radioactivity in each fragment after subtraction of the background is given as a percentage of the total radioactivity. The percentage of total guanines in each fragment was determined from the SV40 nucleotide sequence^{3,4}.

Relaxed circular DNA is a more appropriate control than supercoiled DNA as SV40 chromosomes contain topoisomerase activity and are themselves thought not to be under torsional stress^{26,27}. The free DNA was reacted with ³H-AAAF as described in Table 1. The adduct concentration in SV40 DNA which had been exposed to the carcinogen *in vitro* or *in situ* in the infected cell was estimated from the specific activity of the ³H-AAAF and the radioactivity content of the DNA. SV40 DNA labelled *in vitro* contained an average of about 10 adducts per molecule, whereas SV40 DNA labelled *in situ* contained on average 2 adducts per molecule.

AAAF-modified SV40 DNA was digested with a mixture of the restriction endonucleases *HaeIII* and *KpnI*. This combination gives a set of fragments of convenient size which can be separated by electrophoresis in gels of 3% low-gelling-temperature agarose (Fig. 1). In no case was the AAAF-modified DNA seen to give an altered restriction enzyme digestion pattern compared with normal SV40 DNA. Table 1 shows, for each fragment, its proportions of the total guanine content (calculated from the nucleotide sequence^{3,4}) and of the incorporated radioactivity after *in vivo* or *in vitro* labelling. The amount of radioactivity in fragments labelled as free DNA followed the guanine content; when the specific activity was determined as a function of guanine content (Fig. 2b), all fragments were labelled equally within experimental error. This result is expected if no region of SV40 DNA shows preferential binding of AAAF because over 95% of AAAF-DNA adducts are formed with guanine residues^{15,16,18}. In contrast, the results obtained with SV40 DNA isolated from ³H-AAAF-treated cells indicated non-random distribution of AAAF-guanine adducts. In four determinations the fragment *HaeIII* F1b/*KpnI* was always labelled to a specific activity 1.5–2 times higher than that expected from the guanine content (Fig. 2a). The adduct concentrations in the restriction fragments from the rest of the SV40 genome were proportional to their guanine content.

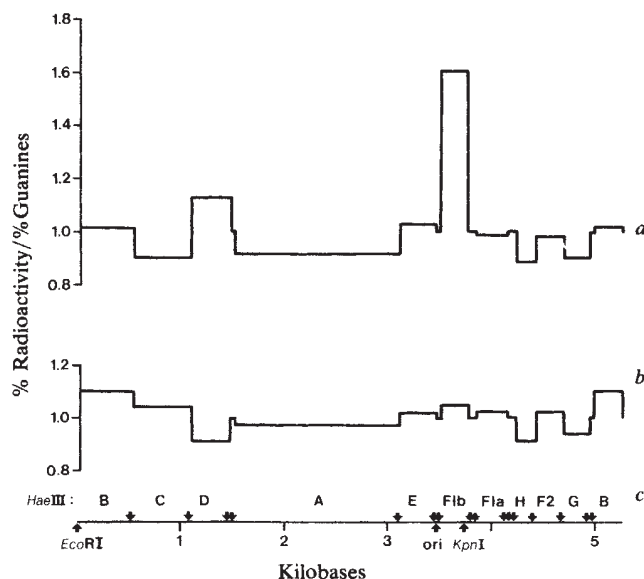


Fig. 2 Preferred reaction site for AAAF on the intracellular SV40 chromosome. The percentage of total guanines present in each SV40 restriction fragment, and the percentage of total ^3H -AAAF radioactivity in each fragment were determined as described in Table 1 legend. The ratio of the percentage of total radioactivity to the percentage of total guanines was calculated for each fragment. *a*, SV40 DNA labelled *in situ* in infected cells; *b*, labelled as free SV40 DNA; *c*, the abscissa is a *Hae*III endonuclease cleavage map of the circular SV40 DNA interrupted at the *Eco*RI site. The map is based on Lebowitz *et al.*³⁸, as is the SV40 nucleotide sequence^{3,4}. DNA fragments too small to be analysed have been assumed to be labelled at the average value for the whole DNA.

Thus in the intracellular SV40 chromosome, a stretch of DNA within ~300 base pairs of the 'late genes' side of the origin of replication (Fig. 2c) is preferentially reactive with AAAF. We do not know whether the region just to the 'early genes' side of the origin is also specifically overlabelled with AAAF, as the restriction fragment containing this region is too small (43 base pairs) to be analysed in these experiments. The next distal fragment to the early side, *Hae*III E (329 base pairs) was labelled normally.

The level found for AAAF modification of the hyper-reactive region is a minimum estimate because we do not know the exact limits of this region on the SV40 genome. If the exposed region is much shorter than the DNA fragment *Hae*III F1b/*Kpn*I which we characterized as containing it, then the level of modification would be correspondingly higher than the value we obtained, which is the average for the whole of that fragment. The level would also be greater if the hyper-reactive region were present in only a fraction of the SV40 chromosomes in the cell.

Why should AAAF react predominantly with a certain region of chromosomal DNA? The higher reactivity of AAAF towards nucleosomal linker compared with core DNA¹⁸⁻²² suggests that the hyper-reactive region of SV40 chromosomes may represent DNA free of protein, or associated with less tightly bound protein. Electron microscopic observation^{11,12} showed a nucleosome-free region on a fraction of isolated SV40 chromosomes; it was unclear to what extent this region already existed *in vivo* or was created by loss of nucleosomes during isolation of nuclei or of the viral chromosomes^{11,12}. From our results, it is likely that a nucleosome-free stretch of DNA exists on SV40 chromosomes also *in situ* in the infected cell. The reaction of AAAF with DNA is also known to favour denatured DNA over native²⁸; it is therefore possible that the hyper-reactive region has an unusual secondary or tertiary structure (reviewed in ref. 29). This region is not, however, simply the most readily denatured part of SV40 as detected by electron microscopy or *S*₁ nuclease sensitivity^{30,31}. It also remains poss-

ible that AAAF induces an alteration of chromosomal structure in the SV40 origin region. However, modification of DNA by AAAF has been reported not to interfere with the *in vitro* reconstitution of apparently normal nucleosomes³².

The region of the SV40 chromosome which reacts preferentially with AAAF contains regulatory sequences controlling the origin of replication of DNA³³, T-antigen binding sites³⁴ and likely promoters for late transcription^{35,36}. SV40 uses cellular enzymes for its replication⁵, and SV40 chromatin probably resembles certain regions of cellular chromatin in structure³⁷. Therefore, regulatory sequences of some cellular genes may likewise be particularly susceptible to modification by AAAF and similar carcinogens.

We thank B. Hirt for advice and criticism of the manuscript, and K. Nyfeler and B. Bentele for technical assistance. This work was supported by grants 3.299.78 and 3.305.78 from the Fonds National Suisse de la Recherche Scientifique.

Received 9 December 1980; accepted 9 March 1981.

1. Sarma, D. S. R., Rajalakshmi, S. & Farber, E. in *Cancer* Vol. 1 (ed. Becker, F. F.) 235-287 (Plenum, New York, 1975).
2. Neidle, S. *Nature* **283**, 135 (1980).
3. Fiers, W. *et al. Nature* **273**, 113-120 (1978).
4. Reddy, V. B. *et al. Science* **200**, 494-502 (1978).
5. Tooze, J. (ed.) *DNA Tumor Viruses* (Cold Spring Harbor Laboratory, New York, 1980).
6. Bellard, M., Oudet, P., Germond, J.-E. & Chambon, P. *Eur. J. Biochem.* **70**, 543-553 (1976).
7. Cremisi, C., Pignatti, P. & Yaniv, M. *Biochem. biophys. Res. Commun.* **73**, 548-554 (1976).
8. Ponder, B. A. J. & Crawford, L. V. *Cell* **11**, 35-49 (1977).
9. Scott, W. A. & Wigmore, D. J. *Cell* **15**, 1511-1518 (1978).
10. Varshavsky, A. J., Sundin, O. & Bohn, M. *Cell* **16**, 453-466 (1979).
11. Saragosti, S., Moyné, G. & Yaniv, M. *Cell* **20**, 65-75 (1980).
12. Jakobovits, E., Bratosin, S. & Aloni, Y. *Nature* **285**, 263-265 (1980).
13. Beard, P. *Cell* **15**, 955-967 (1978).
14. Mathis, D. J., Oudet, P., Wasylyk, B. & Chambon, P. *Nucleic Acids Res.* **5**, 3523-3547 (1978).
15. Kriek, E. *Biochim. biophys. Acta* **355**, 177-203 (1974).
16. Kriek, E., Miller, J. A., Juhl, U. & Miller, E. C. *Biochemistry* **6**, 177-182 (1967).
17. Amacher, D. E. & Lieberman, M. W. *Biochem. biophys. Res. Commun.* **74**, 285-290 (1977).
18. Kaneko, M. & Cerutti, P. A. *Cancer Res.* (in the press).
19. Cech, T. & Pardue, M. *Cell* **11**, 631-640 (1977).
20. Wieseahn, G., Hyde, J. & Hearst, J. *Biochemistry* **16**, 925-932 (1977).
21. Jahn, C. & Litman, G. *Biochemistry* **18**, 1442-1449 (1979).
22. Metzger, G., Wilhelm, F. & Wilhelm, M. *Biochem. biophys. Res. Commun.* **75**, 703-710 (1977).
23. Hirt, B. *J. molec. Biol.* **26**, 365-369 (1967).
24. Radloff, F., Bauer, W. & Vinograd, J. *Proc. natn. Acad. Sci. U.S.A.* **57**, 1514-1521 (1967).
25. Champoux, J. J. & McConaughy, B. L. *Biochemistry* **15**, 4638-4642 (1976).
26. Sen, A. & Levine, A. J. *Nature* **249**, 343-344 (1974).
27. Young, L. S. & Champoux, J. J. *Nucleic Acids Res.* **5**, 623-635 (1978).
28. Levine, A. F., Fink, L. M., Weinstein, I. B. & Grunberger, D. *Cancer Res.* **34**, 319-327 (1974).
29. Davies, D. R. & Zimmerman, S. *Nature* **283**, 11-12 (1980).
30. Beard, P., Morrow, J. F. & Berg, P. *J. Virol.* **12**, 1303-1313 (1973).
31. Mulder, C. & Delius, H. *Proc. natn. Acad. Sci. U.S.A.* **69**, 3215-3219 (1972).
32. Yamasaki, H., Leffler, S. & Weinstein, I. B. *Cancer Res.* **37**, 684-691 (1977).
33. Dana, K. J. & Nathans, D. *Proc. natn. Acad. Sci. U.S.A.* **69**, 3097-3100 (1972).
34. Tjian, R. *Cell* **13**, 165-179 (1978).
35. Ghosh, P. K., Reddy, V. B., Swinscoe, J., Lebowitz, P. & Weissman, S. *J. molec. Biol.* **126**, 813-846 (1978).
36. Handa, H., Kaufman, R. J., Manley, J., Gefter, M. & Sharp, P. A. *J. biol. Chem.* **256**, 478-482 (1981).
37. Shelton, E. R., Wassarman, P. M. & DePamphilis, M. L. *J. molec. Biol.* **125**, 491-514 (1978).
38. Lebowitz, P., Siegel, W. & Sklar, J. *J. molec. Biol.* **88**, 105-123 (1974).

Location of domains in globular proteins

Alexander A. Rashin

Institute of Biological Physics, Academy of Sciences of the USSR, 142292 Pushchino, Moscow Region, USSR

Although it has become widely accepted that domains are the basic units of structure, function and evolution in proteins¹ and it is thought that proteins with complex functions evolve by fusion of genes coding for individual domains^{2,3}, the domains are not uniformly defined. Most commonly, domains are simply the compact and more or less loosely connected areas^{1,4} apparent from a visual inspection of protein models; to avoid subjectivity and ambiguities inherent in visual inspection, certain computer algorithms for location of these 'structural' domains have

recently been proposed⁵⁻⁷. An alternative interpretation is that domains are stable protein fragments found in biochemical experiments⁸⁻¹⁷. I regard them as 'globular fragments' which may refold autonomously and carry specific functions¹, and I propose here a method for location of these globular fragments based on surface area¹⁸ measurements. Applied to several proteins the globular fragments found often coincide with structural domains or are contained within them. In particular the globular fragments found in globins correlate with the two structural domains proposed previously^{4,19}, and do not correlate with the three coding sequences separated by introns in the haemoglobin genes²⁰⁻²².

It has been shown that the surface area buried in a native fold is proportional to the gain in free energy of dehydration and hydrogen bond formation due to folding^{23,24}. As surface area is proportional to the number of interatomic contacts²⁵, it may also serve as a crude estimate of van der Waals' interactions, which are very sensitive to errors in interatomic distances and therefore are usually averaged by counting the number of contacts. These findings are used here to define globular fragments.

A globular fragment may be defined as a segment of a protein chain which, when the rest of the protein is removed, buries a maximum of surface area within its fold. As the contribution of the free energy (estimated as $24 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ of buried surface area²⁴) to its stability is maximal, such a fragment is more likely than others to be autonomously stable.

The surface area buried in stable globular proteins follows the law²⁴

$$B_G = 0.859MW + 0.774MW^2 \times 10^{-5} \\ = 94.49L + 93.654L^2 \times 10^{-3} \quad (1)$$

which relates buried areas, B_G (in $\text{\AA}^2$), to the molecular weight, MW, or the number of residues, L ($MW \approx 110L$) in proteins of 50–320 residues. Non-globular structures have smaller buried areas (B) than those given by equation (1), the ratio B/B_G giving a measure of their globularity. Thus, we may calculate the buried surface areas of all chain fragments in a protein, removing the

rest of the structure but assuming that the native conformation is retained, and analyse their B/B_G values. A maximum of B/B_G close to unity indicates the presence of a globular fragment.

Buried surface areas may be calculated from atomic coordinates¹⁸. The use of the simplified geometry^{25,26} and the fast algorithm⁷ allow such calculations to be made with efficiency equal to those for interatomic distances (see refs 5, 27). The buried areas thus determined are within a few per cent of those calculated using the all-atom algorithm²⁴. Figure 1a shows, for the haemoglobin β chain, a significant difference in buried areas between globular and non-globular fragments of the same length. Fragment 19–68 has the buried area $B = 4,800 \text{ \AA}^2$ and $B/B_G = 0.97$, which indicates its globularity. The fragments of 50 residues beginning at residues 1–10 and 33–55 have $B = 3,500 \pm 250 \text{ \AA}^2$ and $B/B_G = 0.71 \pm 0.05$ —these are non-globular. It has not been determined whether equation (1) is valid for $L < 50$. However, I found that conversion of the difference $B_G - B$, calculated for a number of α helices, to the equivalent free energy yields values only a few kcal mol⁻¹ higher than the experimental stability of an α helix. Thus I used equation (1) for all chain lengths ($L \leq 320$), bearing in mind that it is less justified for $L < 50$. Figure 1b shows a typical B/B_G plot.

A summary of my results (Table 1) shows that globular fragments found here often coincide with structural domains or are contained within them. Ribonuclease is a notable exception: S protein (residues 21–124), found to be a globular fragment, was never listed previously among structural domains⁴⁻⁷. Table 1 also shows that my locations of globular fragments agree well with those determined experimentally. Note that location of the two independent globular fragments within the previously proposed entire discontinuous domain in hen lysozyme agrees with significant mutual displacements of these fragments on denaturation shown by X-ray crystallography¹⁴. A relatively low value of B/B_G for the C-terminal globular fragment of elastase is an artefact due to the unusually large number of small residues in this fragment—this means that the expression $MW/110L$ in equation (1) is invalid. The use of the dependence on MW in equation (1) yields $B/B_G = 0.98$ for this fragment, indicating

Table 1 Structural domains and globular fragments

Protein	Structural domains	Globular fragments		
		Calculated limits	B/B_G	Experiment
Pancreatic trypsin inhibitor	1–58*	1–55	1.10	
Immunoglobulin light chain	1–109†	4–109	0.96	Two globular fragments similar to domains (refs 8–11)
	110–214†	110–209	0.96	
	No continuous domain*	23–124	0.99	
Ribonuclease S	(1–38) + (101–129)*	8–29‡	1.05	Globular fragment in 21–124 (refs 12–13)
Hen lysozyme	39–87	39–98	0.97	
		105–116‡	0.97	
T ₄ lysozyme	1–74§	13–67	1.01	Two independently melting fragments (ref. 8)
	75–164§	78–154	1.05	
	10–111*	17–100	1.11	
Papain	113–207*	112–208	1.03	126–245 (ref. 16)
	27–127*	27–120	1.07	
	128–230*	130–229	0.93	
Thermolysin	1–157†	4–151	1.09	206–316 (ref. 17)
	158–316†	203–316	1.06	
	1–85*	21–70	1.03	
Myoglobin	86–153*	86–152	0.99	
	1–74*	20–63	0.91	
	75–141*	80–140	0.94	
Haemoglobin α chain	1–79*	19–68	0.97	
	80–146*	81–145	0.99	

Globular fragments are detected as maxima of B/B_G , where B is the buried surface area of a fragment and B_G is given by equation (1). If there are several maxima of B/B_G within the same chain region, only one fragment corresponding to the highest B/B_G value is listed.

$B = B_T - B_N$, where B_T is the total inaccessible area of all the spheres representing residues^{25,26} in a given structure and B_N is the area of the spheres made inaccessible by their covalent neighbours along the chain in the same structure. Coordinates used in the calculations are from Brookhaven Protein Data Bank²⁸.

* Taken from ref. 4, haemoglobin chains are aligned to myoglobin. † Taken from ref. 35. ‡ Globular fragments much shorter than 50 residues (see text). § Located from stereodrawing in ref. 36. || Low value due to an artefact; equals 0.98 when corrected (see text).

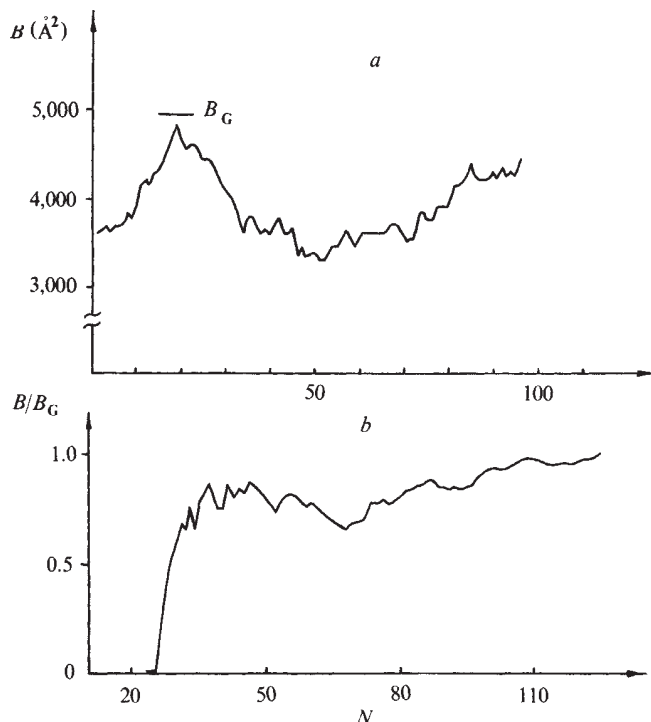


Fig. 1 *a*, The buried surface area B of each segment of 50 residues, beginning at residue N , is calculated for human β chain haemoglobin, assuming that the isolated segment retains its native conformation. The value B_G expected from equation (1) is almost reached by the segment beginning at $N = 19$. *b*, The ratios of the buried area values B of all segments of ribonuclease, beginning at residue 23 and ending at residue N , to the corresponding values B_G expected from equation (1). Two maxima with B/B_G values close to unity at $N = 110$ and $N = 124$ can be seen, the latter being slightly higher.

that this use may be preferable. B/B_G values for globin fragments calculated here using the simplified geometry corresponding to the time-averaged positions of the side chains^{25,26} are higher than those calculated with the all-atom algorithm^{18,24} (or the simplified geometry fitted to the actual positions of the side chains²⁹). This effect of the change in side chain positions agrees with the necessity for small changes of the structure to gain more stability found in the model-building studies of globin fragments²⁹. The lowest B/B_G value listed (which may be supposed to correspond to the lowest stability) is that for the N-terminal globular fragment of haemoglobin α chain. This reflects the lack of D helix in the α chain, which is the major difference between the latter and the β chain or myoglobin, and correlates with the drastically different conformation of the apo- α chain compared with the apo- β chain or the apomyoglobin³⁰.

It was recently proposed, by analogy with immunoglobulins, that translation products of the three continuous sequences or exons in haemoglobin genes correspond to domains^{19–21}. Exon products for the β chain are fragments 1–30, 31–104 and 105–146. Table 1 shows that they correspond to neither structural domains^{4,19} nor globular fragments. Exon products have low B/B_G values of 0.68, 0.85 and 0.81, indicating their non-globularity. They do not seem to correspond to functional domains as the latter are conventionally viewed as being composed of few structural domains (or globular fragments)¹. However, it was shown that fragment 31–104 binds cyanomethaemoglobin but lacks any detectable structure even when bound to haem³¹. Alternatively, it has been proposed³² that the function of each globular fragment in myoglobin is recognition of the haem. It may be speculated that residues 31–63, part of the N-terminal globular fragment, possess marginal stability sufficient to recognize the haem. We found that its B/B_G was 0.94 in the β chain, but this was not confirmed by the all-atom

calculation, suggesting some rearrangement of the structure. This distorted structure may have low helicity undetectable by circular dichroism measurements. These results seem to indicate that at present the domain structure of globins cannot be unequivocally correlated with exons; therefore the problem needs further study.

Algorithms using surface area criteria have been successful in prediction of tertiary^{33,34} and quaternary²⁵ structures of proteins and in location of structural domains (ref. 7 and Wodak and Janin, personal communication). I have shown here that they detect the presence within globular proteins of fragments that may be autonomously stable. Such fragments are likely candidates for intermediates in folding. The folding schemes based on surface area measurements can be derived³² and compared with results obtained experimentally.

I thank B. H. Yudman for help in programming; M. Yu. Pavlov for control calculations using the all-atom algorithm; P. L. Privalov for useful discussion, and J. Janin and S. J. Wodak for useful discussions and for drawing my attention to the domain/exon hypothesis.

Received 9 July 1980; accepted 18 February 1981.

- Schulz, G. E. & Schirmer, R. H. *Principles of Protein Structure* Ch. 5.3 (Springer, New York, 1979).
- Edelman, G. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **63**, 78–85 (1969).
- Rossmann, M. G., Moras, D. & Olsen, K. W. *Nature* **250**, 194–199 (1974).
- Wetlaufer, D. B. *Proc. natn. Acad. Sci. U.S.A.* **70**, 697–701 (1973).
- Crippen, G. M. *J. molec. Biol.* **126**, 315–332 (1978).
- Rose, G. D. *J. molec. Biol.* **134**, 447–470 (1979).
- Wodak, S. J. & Janin, J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1736–1740 (1980).
- Privalov, P. L. *Adv. Protein Chem.* **33**, 167–241 (1979).
- Rowe, E. S. & Tanford, C. *Biochemistry* **12**, 4822–4827 (1977).
- Goto, Y., Azuma, T. & Hamaguchi, K. *J. Biochem.* **85**, 1427–1438 (1979).
- Goto, Y. & Hamaguchi, K. *J. Biochem.* **86**, 1433–1441 (1979).
- Labhardt, A. M. & Baldwin, R. J. *J. molec. Biol.* **135**, 245–254 (1979).
- Chavez, L. G. & Scheraga, H. A. *Biochemistry* **19**, 996–1004 (1980).
- Yonath, A., Podjarny, A., Honig, B., Sielecki, A. & Traub, W. *Biochemistry* **16**, 1418–1424 (1977).
- Johnson, E. R., Anderson, W. L. & Wetlaufer, D. B. *J. biol. Chem.* **253**, 3408–3414 (1978).
- Ghelis, C., Tempete-Gailloudet, M. & Yon, J. M. *Biochem. biophys. Res. Commun.* **84**, 31–36 (1978).
- Fontana, A., Vita, C. & Chaiken, I. M. *Hoppe-Seyler's Z. physiol. Chem.* **360**, 1006–1007 (1979).
- Lee, B. K. & Richards, F. M. *J. molec. Biol.* **55**, 379–400 (1971).
- Pitts, O. B. & Rashin, A. A. *Biophys. Chem.* **3**, 1–20 (1975).
- Gilbert, W. *Nature* **271**, 501 (1978).
- Blake, C. C. F. *Nature* **277**, 598 (1979).
- Eaton, W. A. *Nature* **284**, 183–185 (1980).
- Chothia, C. *Nature* **248**, 338–339 (1974).
- Chothia, C. *Nature* **254**, 304–308 (1975).
- Wodak, S. J. & Janin, J. *J. molec. Biol.* **124**, 323–342 (1978).
- Levi, M. J. *J. molec. Biol.* **104**, 59–107 (1976).
- Rossmann, M. G. & Liljas, A. *J. molec. Biol.* **85**, 177–181 (1974).
- Bernstein, F. C. *et al. J. molec. Biol.* **112**, 535–542 (1977).
- Rashin, A. A. in *Biomolecular Structure, Conformation, Function and Evolution* (ed. Srinivasan, R.) 133–149 (Pergamon, Oxford, 1978).
- Yip, Y. K., Waks, M. & Beychok, S. *J. biol. Chem.* **247**, 7237–7244 (1972).
- Craik, C. S., Buchman, S. R. & Beychok, S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1384–1388 (1980).
- Rashin, A. A. *Studia Biophys.* **77**, 177–184 (1979).
- Richmond, T. J. & Richards, F. M. *J. molec. Biol.* **119**, 537–555 (1978).
- Cohen, F. E., Sternberg, M. J. E. & Taylor, W. *Nature* **285**, 378–382 (1980).
- Liljas, A. & Rossmann, M. G. *A. Rev. Biochem.* **43**, 475–507 (1974).
- Matthews, B. W. & Remington, S. J. *Proc. natn. Acad. Sci. U.S.A.* **71**, 4178–4182 (1974).

O₂ binding properties of the product of the central exon of β -globin gene

Charles S. Craik, Steven R. Buchman & Sherman Beychok*

Departments of Biological Sciences and Chemistry, Columbia University, New York 10027, USA

The hypothesis that the exons of eukaryotic structural genes code for functional domains and that the partitioned arrangement of coding information may thus serve to mediate the rapid evolution of new and unique proteins from pre-existing exons^{1–3} is also supported by our recent studies which demonstrate that the product of the central exon of the human β -globin gene is a complete functional domain capable of binding haem tightly and specifically^{4,5}. Moreover, an analysis of the structure/function

* To whom correspondence should be addressed.

changes induced by mutations in different parts of the haemoglobin molecule suggests that each of the three exon-encoded segments is primarily associated with different functions of haemoglobin (for example, haem-binding, haem-haem interaction)⁶. We have now extended our studies to determine whether the central fragment is sufficient for maintenance of a stable complex of ferrous haem with molecular oxygen and, if not, what are the minimal requirements for the expression of this activity. The results of our reconstitution experiments indicate that the isolated central exon peptide is unable to maintain a ferrous haem-dioxygen complex unless the side exon products and the complementary haem-containing subunit are present. A conformational change which accompanies the noncovalent association of fragments may account for the restoration of reversible oxygen binding.

Haemoglobin and myoglobin in their ferrous states reversibly bind oxygen—this makes them unique among haem-binding proteins and different from most ferrous haem compounds, which are rapidly oxidized in the presence of molecular oxygen. They are the only molecules known, other than the biomimetic model systems (for example, haem-polystyrene matrices⁷ and 'picket fence' porphyrins⁸) that allow O₂ and Fe(II) haem to combine and dissociate freely without concomitant oxidation of the iron. The stability of the oxygenated derivative seems to be achieved in substantial part by accommodation of the haem group in a hydrophobic pocket into which the side chains from the surrounding helices project. These side chains provide the fifth ligand (His F8), modify the binding of the sixth ligand (O₂, CO, CN⁻, etc.) and lock the haem into a specific spatial arrangement with respect to the rest of the protein^{9,10}.

The distinctive spectral characteristics of various ligand derivatives of haemoglobin have been well documented and provide a sensitive means of evaluating the environment and ligation state of the prosthetic group in the protein¹¹. The

method used to investigate the oxygen binding properties of the globin central fragment (see Fig. 1) was based on spectral analysis; its application to the isolated intact globins, after reconstitution with haem, was used as a control. Figure 2a shows spectra resulting from a typical oxygenation experiment in which the intact β -globin chain (β^0) was combined with haemin dicyanide to form the cyanmet subunit derivative, reduced with sodium dithionite and subsequently oxygenated. Spectra for α -subunits of haemoglobin, native or reconstituted, are almost identical to those presented for β , differing only slightly in the absolute value of the extinction coefficients and having a 1–2 nm difference in position of the absorbance maxima.

To follow oxygen binding to the haem-binding domain, by itself and in combination with other components of haemoglobin, the α - or β -globin chain was first proteolytically cleaved with the arginine-specific protease into the fragments corresponding to the exon-encoded regions of the protein (Fig. 1). The visible and Soret spectra of these samples after their reconstitution with haemin dicyanide were then obtained along with their reduced and oxygenated derivatives. Comparison of Fig. 2b with the intact control shown in Fig. 2a reveals almost identical spectra for haem-reconstituted β -globin and β -globin

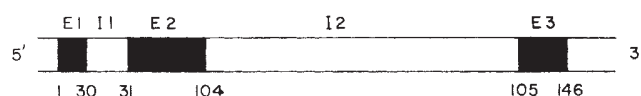


Fig. 1 Diagrammatic representation of the β -globin intron/exon structure.

The black boxes E1, E2 and E3 represent expressed DNA sequences (exons); the white boxes I1 and I2 represent intervening sequences (introns)¹⁶. The open-boxes at the 5' and 3' ends are untranslated regions of indeterminate length. The numbers below the exons denote amino acid sequence positions of the β -globin. E2 represents the central exon that codes for the haem-binding polypeptide consisting of amino acids 31–104. The 5' $\rightarrow$ 3' orientation is indicated. A similar arrangement exists for the α -globin gene^{17,18}. The small intron (I1) for α -globin occurs between codons 31 and 32 whereas the large intron (I2) interrupts the coding sequence at 99 and 100. When the amino acid sequences of α - and β -globin are aligned to maximize functional homology¹⁹, their intron/exon junctions coincide; α -globin residue 31 is an arginine but residue 99 is a lysine, thus limited cleavage products were separated by a modification of the procedure used elsewhere⁴ and involved gel permeation high performance liquid chromatography in 6 M urea, 0.2 M formic acid, using a recycle purification mode²⁰. This procedure is much more rapid and gives higher purity than molecular sieve chromatography in 9% formic acid.

digest. However, whereas reduction with dithionite produced a deoxyhaemoglobin spectrum for the reconstituted β -globin, a haemochromogen spectrum (hexacoordinate ferrous haem iron axially bound to two histidines) was obtained with the reconstituted digest and with the isolated central fragment. Subsequent removal of the dithionite produced an oxyhaemoglobin spectrum for the undigested samples; in contrast, the spectrum of the digested sample or the central fragment was that of a haemichromogen (hexacoordinate ferric haem with two histidines serving as the axial ligands), diagnostic of a haem complex that does not reversibly bind oxygen. Similar results were found for α -globin.

Comparable experiments were carried out on a fully reconstituted sample of β -globin digest (β_{rec}^0), prepared by mixing equimolar concentrations of the various arginyl peptides then combining the mixture with haemin dicyanide. Although the mixture showed a substantial Soret extinction coefficient, haemochromogen and eventual haemichromogen spectra resulted from dithionite addition and its subsequent removal. However, when the complementary haem-containing subunit, α^h , was combined with β_{rec}^0 or the β_{digest}^0 before haem addition, the typical oxy spectrum presented in Fig. 2c resulted. By subtracting the contribution to this oxy spectrum due to the native α^h , it was established that 86% of the β_{digest}^0 produced an oxy spectrum, with the other 14% haemichromogen. The β_{rec}^0

Table 1 Far UV circular dichroism and secondary structure characteristics

Protein or peptide	$[\theta]_{222\text{ nm}}$ (deg cm ² per dmol)	Calculated % α helix*	Calculated no. of helical residues†
β^h	21,600	73	106
$\beta^0 + h$	21,600‡	73	106
$\beta_{\text{digest}}^0 + \alpha^h + h$	18,200		
	(14,800)§	52§	76
$\beta_{\text{rec}}^0 + \alpha^h + h$	17,500		
	(13,400)§	47§	69
β^0	10,600	39	57
$\beta_{\text{digest}}^0 + h$	10,500	38	55
$\beta_{\text{rec}}^0 + h$	9,500	35	51
β_{digest}^0	7,300	28	41
β_{rec}^0	7,500	29	42
$\beta_{\text{cf}}^0 + \alpha^h + h$	13,700		
	(5,800)§	23§	17
β_{cf}^0	5,800	23	17
α^h	22,000	78	110
$\alpha^0 + h$	21,500‡	73	103
$\alpha_{\text{digest}}^0 + \beta^h + h$	18,900		
	(16,200)§	57§	79
$\alpha_{\text{digest}}^0 + h$	14,000	49	69
α^0	11,000	40	56
α_{digest}^0	7,200	28	39

α^0 or β^0 indicates that the chain has no haem attached; α^h or β^h indicates that the chain has bound haem. The subscript 'digest' refers to unfractionated clostripain-treated globin; the subscript 'rec' refers to a fully reconstituted sample prepared by mixing together equimolar concentrations of the various arginyl peptides, and the subscript 'cf' denotes the isolated central fragment of β -globin. $[\theta]_{222}$ values are means of at least three determinations, except for β_{rec}^0 which is the mean of two determinations. The estimated uncertainty in $[\theta]_{222}$ is $\pm 5\%$. In the cases of the β_{rec}^0 samples, the second digit in the calculated α helix may not be significant due to uncertainty in concentration determination.

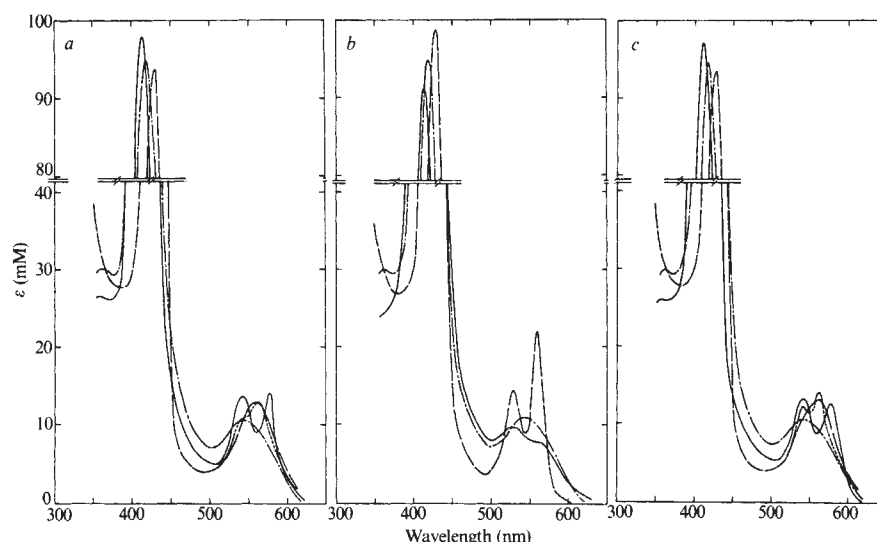
* Calculated according to Chen *et al.*²¹.

† Calculated by multiplying the fraction of helical residues of the third column by 146 and 141 for α - and β -globins, respectively, and 74 for β_{cf}^0 .

‡ Reconstitution of α - and β -globins with haemin dicyanide to give the same $[\theta]_{222}$ requires dialysis into 0.1 M potassium phosphate buffer, 1 mM EDTA, pH 7.4.

§ These values refer to the β components after subtracting the contribution made by α^h .

Fig. 2 Spectral properties of cyanmet-, deoxy- and oxy-derivatives of undigested and digested globins. *a*, Cyanmet-spectrum (—) of an equimolar mixture of β -globin and haemin dicyanide; transient ferrocyanide spectrum (-----) immediately after sodium dithionite addition and the eventual deoxy-spectrum (---); oxy-spectrum (——) after dithionite removal. The transient ferrocyanide spectrum and the deoxy-spectrum are coincident below 500 nm. Solutions of the ferricyanide derivatives showed an absorption spectrum in the visible with a broad band centred at 540 nm and a barely discernible shoulder at 560 nm. The Soret band lies near 420 nm, and has an extinction coefficient about 10-fold greater than that at 540 nm. Ferricyanide derivatives of haemoglobin react directly with dithionite, leading to reduction of the haem iron with ligand still bound, the transient ferrocyanide derivatives then dissociating into cyanide and deoxygenated ferrous haem-protein. The latter transition from ferrocyanide to deoxyferrohaem could be monitored for the separated α - and β -subunits, as the ferrocyanide spectrum, with a maximum at 562 nm and a smaller peak at 541 nm, only gradually coalesced to the deoxyferrohaem spectrum, characterized by a single, broad asymmetric band in the visible region with a maximum at 556 nm and a Soret band in the near UV at ~ 430 nm. After removal of dithionite by dilution, dialysis or gel filtration, and exposure of the haem-protein solution to oxygen, the characteristic oxyferrohaem spectrum appears. The two visible absorption bands are quite distinct from each other, with one peaking at 577 nm and a broader one of slightly lower intensity peaking at 541 nm. The Soret band is at 415 nm. *b*, Cyanmet-spectrum (—) of an equimolar mixture of 60-min clostripain digested β -globin and haemin dicyanide; haemochromogen spectrum (---) due to dithionite addition and a haemichromogen spectrum (——) after dithionite removal. For α -globin, this spectrum also contains a 9.3% oxy-component presumably due to the small amount of undigested globin present. *c*, Cyanmet-spectrum (—) of an equimolar mixture of 60-min clostripain digested β -globin, the complementary haem-containing α -subunit and haemin dicyanide; transient ferrocyanide spectrum (-----) immediately after dithionite addition and the eventual deoxy-spectrum (---); oxy-spectrum (——) after dithionite removal. The transient ferrocyanide spectrum and the deoxy-spectrum are coincident below 500 nm. Measurements were made at 4 °C in 20 mM potassium phosphate, 1 mM EDTA buffer, pH 5.7 or 0.1 M potassium phosphate buffer, 1 mM EDTA buffer, pH 7.4. Although not essential for α^0 reconstitution with haem, 0.1 M potassium phosphate pH 7.4 buffer was required to develop reversible oxygen binding activity for the haem-reconstituted β -subunit. To reduce undesirable side effects, dithionite reductions were achieved with a 10-fold excess of dithionite in a nitrogen atmosphere. Dithionite removal was by gel filtration through a Bio-Gel P-2, dialysis or dilution.



sample yielded a lower value of 74%. This discrepancy was probably due to irreversible denaturation of the fragment produced by the separation procedures.

No appreciable oxy spectrum was observed when the experiment was repeated with the central fragment alone mixed with α^h . It is therefore clear that the reversible O_2 binding abilities of the central exon haem-binding fragment, incapable of reversibly binding O_2 on the time scale (min) and in the conditions of the above experiments, can be restored by the presence of the side exon fragments and the complementary haem-containing subunit.

To determine whether reversible O_2 binding is correlated with secondary structure, the far UV circular dichroism spectra of the haem-free globins, the fractionated, unfractionated and reconstituted digestion products, the central fragment of β -globin and their haem-reconstituted mixtures were determined; the results are presented in Table 1. The helix contents of both α - and β -globins, which are half those of the haem-containing chains, or lower depending on conditions, were reduced further by digestion with clostripain to the extent of an additional 10% loss in helicity. Both α^0 and β^0 digests, as well as β^0_{rec} , showed some recovery of helix content on recombination with haemin dicyanide: $\sim 20\%$ (or 30 residues) refolding in the case of α^0_{digest} and $\sim 10\%$ in the case of β^0_{digest} and β^0_{rec} . The addition of the complementary haem-containing subunit α^h to β^0_{digest} or β^0_{rec} , and β^h to α^0_{digest} induced some refolding, $\sim 10\%$ in the case of α^0_{digest} and 15% in the case of β^0_{digest} . These changes are smaller in magnitude than those seen when α^h is added to β -globin or β^h is added to α -globin, perhaps reflecting the limitations imposed on refolding by the interruption of covalent bonding in the linear sequence of residues of the native helices. Neither addition of haemin dicyanide or α^h affected the circular dichroism of the isolated β^0 central fragment.

The loss of structure of the globins on clostripain cleavage, as seen from the circular dichroism measurements, seems to be sufficient for coordination of the distal histidine with the haem iron, thus preventing oxygen binding and resulting in haemichromogen formation. Once the ability to restore the native three-dimensional structure of the haem-binding cavity has been lost, the maintenance of a stable ferrous complex in the

presence of oxygen is prevented. This is clearly the case for haemoglobin or myoglobin where the stability of the oxygenated derivative is lost on denaturation, linking the integrity of the conformation of the protein to its O_2 -binding capabilities.

The possibility that the structural features defining the haem pocket are induced by haem-binding to the central fragment and can be made more precise when the side fragments are present, even if not covalently joined, has been suggested from Soret extinction coefficient measurements of haem-reconstituted β^0 , β^0_{digest} , β^0_{rec} and the central fragment. However, this was not sufficient to restore secondary structure or reversible oxygen binding to the haem-binding fragment. Only by introducing equivalent amounts of the complementary haem-containing subunit before reconstitution with haem could an oxy-haemoglobin spectrum result. This was accompanied by a 14% increase in helix content (20 residues) which suggest a restructuring of the haem pocket to allow reversible O_2 binding. The phenomenon is reminiscent of the refolding observed in complementation studies on ribonuclease¹² and staphylococcal nuclease¹³.

These results, taken together with our earlier studies, demonstrate that the exon-encoded central fragment has the structural potential of providing a tight and specific binding site for haem, and that the fit is sharpened by the addition of the N- and C-terminal fragments, but that formation of a stable dioxygen complex requires, in the conditions of our experiments, the presence of the side fragments and the complementary subunit.

We are investigating, by stop-flow methods, whether transient reversible oxygen binding in the absence of the complementary subunit occurs, to elucidate the mechanism by which the distal histidine becomes irreversibly bound to the iron. Obviously, the requirement of a complementary subunit cannot be general, as shown by the reversible oxygen binding of the α subunit and myoglobin, both monomeric. In this case, the added subunit served to overcome the tendency for the distal histidine to be bound to haem in the deoxy state. However, it is precisely a difference in bonding tendency of the distal histidine that distinguishes the globins from the *b*-type cytochromes¹¹. The central exon product, isolated or in noncovalent association with

the side peptide fragments, accordingly has spectral characteristics reminiscent of these cytochromes, and undergoes a similar reduction-reoxidation cycle with dithionite and molecular oxygen.

The finding of a haem-binding (or, more generally, porphyrin-binding) domain encoded by an exon that can combine with other gene elements to generate a multiplicity of haem-binding proteins of diverse function, would be very fundamental. Thus we have begun to define the essential features of a haem-binding site and have shown similarity in the nature and spatial distribution of haem contact residues in haemoglobin chains and several *b*-type cytochromes^{5,14,15}. We will provide details of these similarities elsewhere and suggest that the evolution of the globin genes from earlier cytochrome-like ones involved combination of an exon coding for a haem-binding domain with others not derived from cytochrome genes.

This study was supported in part by NSF grant PCM 80-05614. We thank Drs Alex Tzagoloff and Jonathan Greer for critical reading of the manuscript and constructive advice, and Drs Jeremy Luchins and Yvonne Leutzinger for their interest and support. The authors also thank Ms Jill Macoska for preparation of the text, table and figures.

Received 15 January; accepted 2 March 1981.

- Gilbert, W. *Nature* **271**, 501 (1978).
- Blake, C. C. F. *Nature* **277**, 598 (1979).
- Gilbert, W. in *Eucaryotic Gene Regulation* (eds Axel, R., Maniatis, T. & Fox, C. F.) 1-12 (Academic, New York, 1979).
- Craik, C. S., Buchman, S. R. & Beychok, S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1384-1388 (1980).
- Craik, C. S. & Beychok, S. in *Interaction Between Iron and Proteins in Oxygen and Electron Transport* (ed. Chien Ho) (Elsevier North-Holland, New York, in the press).
- Eaton, W. *Nature* **284**, 183-185 (1980).
- Wang, J. H. *Acc. Chem. Res.* **3**, 90-97 (1970).
- Collman, J. P. *Acc. Chem. Res.* **10**, 265-272 (1977).
- Perutz, M. F. *A. Rev. Biochem.* **48**, 327-386 (1979).
- Perutz, M. F. *Scient. Am.* **239**, 92-125 (1978).
- Lemberg, R. & Legge, J. M. *Hematin Compounds and Bile Pigments* (Interscience, New York, 1949).
- Brown, J. E. & Klee, W. A. *Biochemistry* **10**, 470-476 (1971).
- Taniuchi, H., Anfinsen, C. B. & Sodja, A. *Proc. natn. Acad. Sci. U.S.A.* **58**, 1235-1242 (1967).
- Leutzinger, Y. & Beychok, S. *Proc. natn. Acad. Sci. U.S.A.* **78**, 780-784 (1981).
- Nobrega, F. G. & Tzagoloff, A. *J. biol. Chem.* **255**, 9828-9837 (1980).
- Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. *Cell* **15**, 1157-1174 (1978).
- Nishioka, Y. & Leder, P. *Cell* **18**, 875-882 (1979).
- Lauer, J., Shen, C.-K. J. & Maniatis, T. *Cell* **20**, 119-130 (1980).
- Dickerson, R. E. & Geis, I. *The Structure and Action of Proteins*, 52 (W. A. Benjamin, California, 1969).
- Craik, C. S. thesis, Columbia Univ. (1980).
- Chen, Y.-H., Yang, J. T. & Chau, K. H. *Biochemistry* **13**, 3350-3359 (1974).

Correlation of DNA exonic regions with protein structural units in haemoglobin

Mitiko Gö

Department of Biology, Faculty of Science, Kyushu University 33, Fukuoka 812, Japan

The discovery of intervening sequences (introns) in DNA led Gilbert and Tonegawa^{1,2} to suggest that a new protein could have been produced by bringing together certain segments of pre-existing ones. However, Blake³ argued that if DNA was so organized that coding sequences (exons) correspond to structural as well as functional units of proteins, then combinations would be much more likely to yield a stable globular conformation through being 'sums of parts'. In immunoglobulin heavy chain, four separate exons encode four different units⁴, all with distinct functions and three of which have clear domain structures. However, in haemoglobin, which has no obvious

domain structure, no clear conformational characteristics have so far been recognized for the segments encoded by exons. From a close inspection of their conformations by drawing various stereodiagrams and the C^α-C^α distance map, I now propose a conformational characterization of the segments as structural units.

In the mouse α - and β -globin genes, the three coding sequences correspond to amino acid residues 1-31, 32-99 and 100-141, and 1-30, 31-104 and 105-146, respectively⁵⁻⁸. These two subunits of haemoglobin are structural homologues. The central coding sequence has been shown to correspond to the function that interacts with and binds haem^{9,10}. Do exonic regions of DNA correspond to structural units of protein conformation in haemoglobin? It is well known that the globin monomers do not show domain or supersecondary structures¹¹ such as those observed in immunoglobulin or dehydrogenases. A domain can be regarded as a part of a protein which forms a compact, continuous globular region separated from the rest of the molecule^{12,13}. The joints at residues 30-31 (31-32 in the β chain) and 104-105 (99-100 in the β chain) between adjacent segments encoded by different exons are located in the middle of B and G helices, respectively. Thus the exonic regions do not correspond to simple secondary structures. Therefore, the conformations of the segments are not characterized as domains, supersecondary or secondary structures.

The folding structure of the whole β chain and those of its three segments encoded by the exons are shown in Fig. 1a-d. The segment coded by the right exon consists of two helices connected by several residues forming a reverse turn (Fig. 1b). This segment is in slight contact with the haem group. The segment encoded by the central exon compactly encloses the haem group; 14 out of 17 side chains in contact with the haem are part of this segment (Fig. 1c). This correspondence of haem binding to the central coding region has been reported elsewhere^{9,10}. The segment encoded by the left exon is a helical arc and is less compact than the other segments (Fig. 1d)—this segment is not in contact with the haem group. The mutual arrangements of the segments in Fig. 1a are such that they lie, and are thus exposed on, the surface of the molecule. They may thus be regarded as a surface layer of the molecule. There is no core in haemoglobin subunits, that is, the subunits consist solely of the surface layer, which is due to the small size of the subunits. Note also that in Fig. 1a the joints between adjacent segments occur relatively close to the centre of the molecule. In fact the residue accessibility to water molecules is extremely small at these joints¹⁴.

Protein folding can be presented on a two-dimensional plane by the so-called C^α-C^α distance map¹⁵ first used by Rossmann and Liljas¹³ to locate folding domains. By careful inspection of various patterns on the distance maps, Kuntz¹⁶ showed that the square and trapezoidal patterns represented the simplest and smallest units of the tertiary structure consisting of the three segments that are situated close together in space. Folding of the haemoglobin β chain is shown on this distance map in Fig. 2. The pairs of residues separated by >27 Å are shown by dark shading: these regions are grouped into seven major continuous areas. In Fig. 2 solid lines are drawn at the joints between the polypeptide segments encoded by different exons, that is, between residues 30 and 31 and between 104 and 105. It is remarkable that these lines scarcely cross the dark regions on the map.

This fact together with the characteristics observed in the stereodiagrams leads to a clear characterization of the conformations of the segments encoded by the exons. To describe this characterization, consider the following problem. Given the three-dimensional structure of each subunit of haemoglobin as shown in Fig. 1a, when cut into three or four segments consisting of 30-50 residues, in such a way that they are as little extended as possible, it is obvious that all conformations of such segments cannot be very compact. One or two must be rather extended. A possible algorithm for solving this problem in such small one-layer proteins as haemoglobin subunits uses pairs of residues

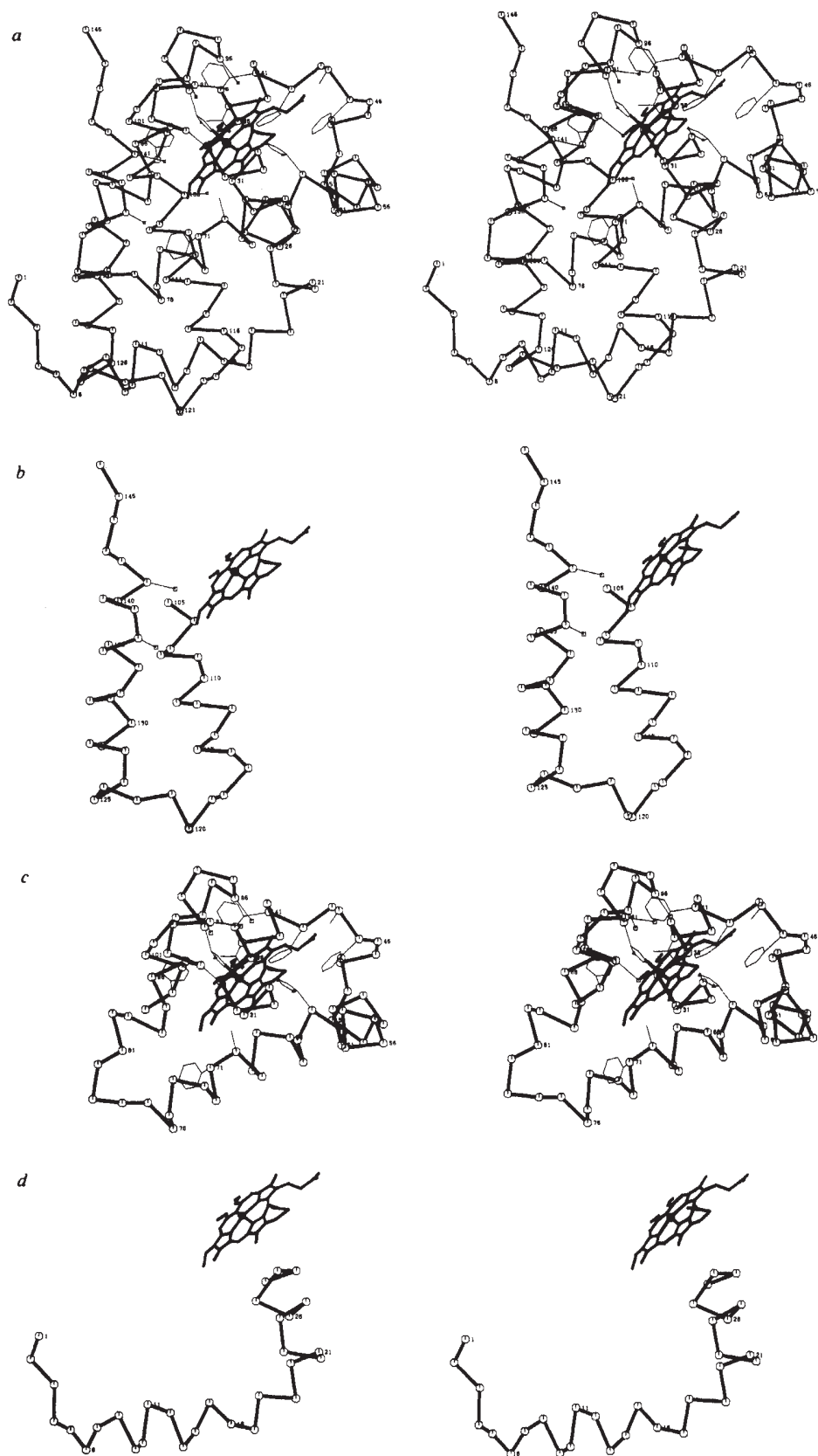


Fig. 1 Stereodiagrams of the whole β chain (a) and of the segments corresponding to the right (b), central (c) and left exons (d). The α -carbon backbone is represented. The haem group is shown with each segment to give the relative three-dimensional position. The side chains in contact with the haem group²⁰ (located within ~ 4 Å) are drawn symbolically.

which are very far apart in space and cut the polypeptide chain in such a way that pairs of residues that are far apart are not included within each individual segment. This is exactly the situation seen in Fig. 2 if we introduce one more cut somewhere between residues 66 and 71. A set of four segments, F1-4, presents a solution to our problem. F1, F2+F3 and F4 are

encoded by exons. From inspection of Fig. 1a the central segment itself is seen to be very little extended. Thus, conformations of segments encoded by exons are characterized by the fact that they are the least extended of all possible choices of cuts in the polypeptide chain consisting of three or four segments of 30-50 residues. Pairs of residues which are far apart usually

occur at two opposite sides of a globular structure. Therefore, cuts which are introduced to separate them tend to occur close to the centre of the molecule (Fig. 1a).

The structural units can be defined by the present algorithm for proteins whose X-ray data are available. Lysozyme is one of these proteins—its gene sequence has been reported by Jung *et al.*¹⁷. It will be reported elsewhere that three out of four junctions between adjacent structural units of lysozyme determined by the present algorithm correspond to the exon-intron junctions.

Conformation of each segment would not be stable when isolated¹⁸ but the ability to assume 'the native format', when incorporated into a globular structure together with other segments, is kept at a maximum by 'being least extended' among all possible choices. In a small one-layer protein molecule, one side of the conformation of each segment is exposed to the surface and therefore should be mainly polar; the other side should be mainly nonpolar. The probability of forming a new protein from segments of old proteins is expected to be fairly high, when segments have the characters described above. The conformational characteristics now recognized for the segments encoded by the exons in haemoglobin subunits make it more likely that they have a structural, as well as a functional role in the process of protein evolution.

From the structural point of view it is more natural that the structural units F2 and F3 are encoded by separate exons. However, no intervening sequence has been found in the DNA at a position corresponding to the F2-F3 junction. One possible explanation for this absence is that selection pressure eliminated the intervening sequence that was present in an ancestral pro-

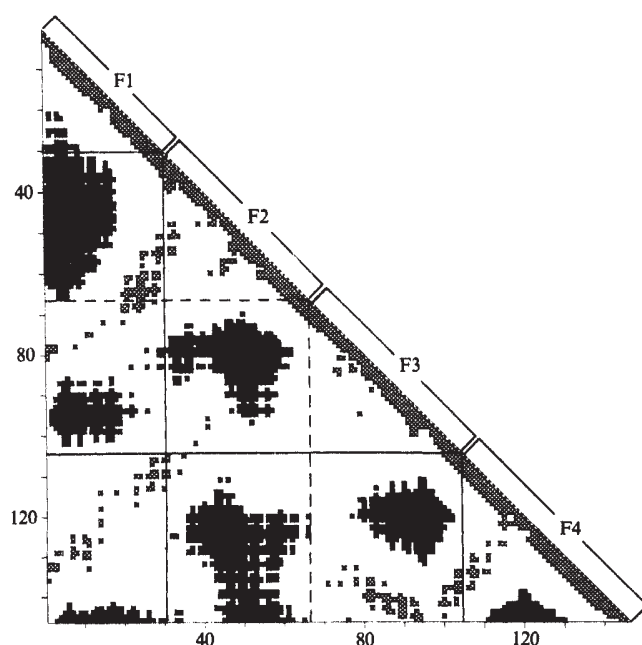


Fig. 2 Distance map of haemoglobin β chain. The distances between the i th and j th α -carbon atoms are plotted in a triangular space, in which both ordinate and abscissa are the residue sequence number. The grey region represents the pairs of C^α atoms within 9 Å, the open area the pairs whose distances are 9–27 Å, and the dark region the pairs separated by >27 Å. α -Helices are represented by a series of interactions along the diagonal. The structural units F1–F4 are defined by considering the exonic regions, this map and the amino acid sequences near the joints of the units. The units F1, F2+F3 and F4 are encoded by the three exons, respectively. The atomic X-ray data used are of horse deoxyhaemoglobin (provided by the Protein Data Bank at Brookhaven National Laboratory).

Structural unit	-F1- F2- F3- F4-									
α chain										
Residue no.	---31	32	33---	61	62	63---	99	100	101--	
Common	Arg			Lys			Lys		Leu	
Major		Met	Phe		Val	Ala			Leu	
β chain										
Residue no.	---30	31	32---	66	67	68---	104	105	106--	
Common	Arg	Leu	Leu	Lys	Val			Leu	Leu	
Major							Leu	Arg		

Fig. 3 The amino acid residues that are common and major in various species²¹ at the junctions of the structural units in haemoglobin α and β chains.

tein¹⁹. This is likely because the structural units F2 and F3, when put together, fix the haem group in its functional position, and because unnecessary exon shuffling promoted by presence of the intron would be harmful. In connection with this, note that an interesting feature of the amino acid sequences is observed when rather arbitrary cuts between F2 and F3 are introduced between residues 61 and 62 in the α chain and between 66 and 67 in the β chain (Fig. 3). The last residues of F1 in both chains, and of F3 in the α chain are common in many species—these are arginine or lysine. Arginine is the main residue at the last position of F3 in β chain—lysine is less common. The last residue of the F2 in both chains is commonly lysine. Furthermore, the first two residues of F2 and F4 are hydrophobic, which is again the case for the first two residues of F3. Although it is not known why lysine and arginine, followed by hydrophobic residues, are found at the joints between the segments encoded by exons, the common character of the amino acid residues at the three joints of the segments supports the possibility of the presence in, and subsequent elimination from the ancestral globin, of an intron between DNA sequences encoding the structural units F2 and F3.

I thank Mr H. Mizuno for the use of the computer programs for the distance map and stereodiagram, Professor H. Matsuda for his encouragement during the work, Dr N. Gō for critical reading of the manuscript and various discussions, and Professor G. Schütz for providing the lysozyme exon-intron structures before publication. Part of this work was supported by grants-in-aid from the Ministry of Education, Japan.

Received 25 September 1980; accepted 16 February 1981.

- Gilbert, W. *Nature* **271**, 501 (1978).
- Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1485–1489 (1978).
- Blake, C. C. F. *Nature* **273**, 267 (1978).
- Sakano, H. *et al. Nature* **277**, 627–633 (1979).
- Leder, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 6187–6191 (1978).
- Nishioka, Y. & Leder, P. *Cell* **18**, 875–882 (1979).
- Konkel, D. A., Tilghman, S. M. & Leder, P. *Cell* **15**, 1125–1132 (1978).
- Konkel, D. A., Maizel, J. V. Jr & Leder, P. *Cell* **18**, 865–873 (1979).
- Craik, C. S., Buchman, S. R. & Beychok, S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1384–1388 (1980).
- Eaton, W. A. *Nature* **284**, 183–185 (1980).
- Rao, S. T. & Rossmann, M. G. *J. molec. Biol.* **76**, 241–246 (1973).
- Wetlauffer, D. B. *Proc. natn. Acad. Sci. U.S.A.* **70**, 697–701 (1973).
- Rossmann, M. G. & Liljas, A. *J. molec. Biol.* **85**, 177–181 (1974).
- Gō, M. & Miyazawa, S. *Int. J. Peptide Protein Res.* **15**, 211–224 (1980).
- Nishikawa, K., Ooi, T., Isogai, Y. & Saito, N. *J. phys. Soc. Japan* **32**, 1331–1337 (1974).
- Kuntz, I. D. *J. Am. chem. Soc.* **97**, 4362–4366 (1975).
- Jung, A., Sippel, A. E., Grez, M. & Schütz, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5759–5763 (1980).
- Anfinsen, C. B. & Scheraga, H. A. *Adv. Protein Chem.* **29**, 205–300 (1975).
- Doolittle, W. F. *Nature* **272**, 581–582 (1978).
- Perutz, M. F., Muirhead, H., Cox, J. M. & Goaman, L. C. *G. Nature* **219**, 131–139 (1968).
- Dayhoff, M. O., Hunt, L. T., McLaughlin, P. J. & Barker, W. C. *Atlas of Protein Sequence and Structure* Vol. 5 (ed. Dayhoff, M. O.) D370–D371 (National Biomedical Foundation, Maryland, 1972).

BOOK REVIEWS

Not by bones alone: "The Making of Mankind"

Barry Cox

The Making of Mankind. By Richard E. Leakey. Pp. 256. ISBN 0-7181-1931-2. (Michael Joseph: 1981.) £9.95. Dutton will publish the book in the USA in September, price \$25.75. The first programme in the seven-part television series was shown on BBC 2 on 4th May and the remainder will appear on successive Mondays at 9.50 pm.

IN the now common double-barrelled "Let 'em have it in the eyes and ears together" approach, the publication of Richard Leakey's book has accompanied the beginning of the seven-part BBC TV series of the same title — *The Making of Mankind*. After seeing a preview of the first and third programmes, reading the book allows one to make a reasonable guess at the nature of the complete series. I think, on the whole, that Richard Leakey deserves a high mark for balance, content and style.

The linkman of a series of programmes of this kind has an immense opportunity. Taken wherever there is something worth filming, with researchers to help in identifying and gathering relevant material, he can usefully indulge that desire to read and think beyond the boundaries of his own specialist research that lies in many scientists. Leakey has grasped this opportunity with both hands; he quotes Alan Walker's words: "If all we were to learn from prehistory was a list of species in time and space, then we should give up". *The Making of Mankind* is not, therefore, merely an anatomical/stratigraphical account of how one type of leg-bone evolved into another, but is an attempt to weave together all the different threads of evidence that may help us to understand the pathway that led to man today. This approach also has the advantage for the television viewer that he does not have the uneasy feeling that he must, from week to week, remember what went before or risk misunderstanding later programmes.

The beginning of such a series is always the most difficult to design, since one has to capture the interest of the audience and at the same time give an idea of the scope of the whole and provide some of the inevitably rather unglamorous chunks of basic background. Leakey starts with attractive film of exploration and discovery in East Africa, and then back-tracks to explain the processes of evolution and fossilization, and the geological time-scale. I think it was unnecessary to start with the stock sequence (photographs of volcanic activity — origin of the world — origin of life — metazoa — life on land etc.), and I also cannot see that the current debate over graduated evolution versus punctuated

equilibria throws any light on human evolution. But the animation—explanation of the process of fossilization was extremely clear, and well integrated with a real-life simulation of rain-water revealing a fossil skull. The programme ends with an account of the earliest hominoid fossils, the forest dryopithecines and the woodland ramapithecines, and suggests how climatic change and consequent environmental change may have moulded this stage of human evolution. On the whole, I felt that this programme negotiated the initial problems very well.

The second programme looks at the African discoveries of fossils of *Australopithecus* and early types of *Homo*, and at the evidence that at least two species of hominid were apparently able to co-exist. It therefore includes the problematical Hadar fossils and the differing interpretations of Leakey and Johansen over their nature: do they all belong to one species of *Australopithecus*, or do some belong to an early species of *Homo*? Though Leakey has been accused of giving a one-sided view of the debate, that is certainly not true of this section of the book, while Don Johansen himself presents his own case in the TV programme, so I feel that the accusation is

quite unfair — and the more regrettable for encouraging the layman's view that scientists cannot hold differing theories without becoming personally involved.

It is the third programme that Leakey starts to widen the scope of the series, emphasizing the new experimental approach to the problems of understanding the making and use of flint tools and the social significance of the camp-sites of early man. Here, too, he introduces his main message: that the fundamental social characteristic of man is that he is a *sharer*, his camp-sites being the locations to which food was brought to be shared. This theory, explained by Glynn Isaac, suggests that it was the contribution of both sexes to the food-sharing group (the males scavenging for meat while the females collected plant foods) that provided the basis for the eventual evolutionary superiority of *Homo* over *Australopithecus*. This way of life is presented dramatically by the use of actors made to resemble *Homo erectus* by wearing startlingly life-like masks, and by film of the modern !Kung bushmen who today follow the hunter-gatherer way of life into which this food-sharing habit may have evolved.

In later chapters and programmes, Leakey outlines recent work on the ability

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Face from the past — an actor wearing the *Homo erectus* mask.

of apes to use signs and, by implication, to classify objects and concepts, and on the brain casts of fossil man. (The tendency to suggest that a larger brain *allows* more complex behaviour surely puts the morphological cart before the functional horse?). Neanderthal man leads on to Ice Age art; film cameras were allowed inside Lascaux for the first time in 20 years, and we are promised "staggering and beautiful images of horse, deer, bison and stags". It also seems as though horses and cattle may have been domesticated far earlier than had been thought. Leakey ends with a consideration of the implications of the agricultural revolution.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Richard Leakey at work.

What, then, of the whole? Leakey himself comes over as a sympathetic and relaxed linkman, whose own personality does not obtrude between the viewer and the story, and the TV series is beautifully filmed by Alec Curtis and John Record. The scope of both the series and the book is satisfyingly broad and, as we shall see, there is a message of hope. The book is well written and illustrated, though I could not help feeling that a figure showing the relationships of the hominids to one another in time and their possible evolutionary relationships would have been helpful. Though there is a good index, there is no bibliography of any kind, not even to other popular books.

The culmination of Leakey's message comes in his consideration of the origins of human aggression. He suggests that it was only after the agricultural revolution, about 10,000 years ago, that the possession of settled and accumulated property made its gaining by aggression profitable and its defence therefore necessary. So, Leakey concludes, "Warfare is a social and political response to changed economic circumstances. Humans are essentially cultural creatures capable of responding in many different ways to the same prevailing circumstances". For him, the ills of the modern world are not the ultimate result of the irrevocable crystallization of our social nature during some early man-the-hunter phase of evolution, but are a comparatively recent adaptation — and therefore reversible. Let us hope that history will prove him right. □

Barry Cox is Professor of Zoology at King's College, University of London.

Glimpsed behind the curtain

Vera Rich

The Social Context of Soviet Science. Edited by L.L. Lubrano and S.G. Solomon. Pp.245. ISBN 0-8919-0712-9. (Dawson, Folkestone, Kent/Westview, Boulder, Colorado: 1980.) £10, \$26.25.

ALTHOUGH the average Western scientist may well have numerous contacts with Soviet colleagues, and may himself have visited the Soviet Union, he will not necessarily have an overall view of the structure of Soviet scientific life. This is not for lack of information — as Susan Gross Solomon observes in her introductory essay, the Western study of Soviet science is now entering its fourth decade, with a proliferation of research projects addressed to governmental as well as academic audiences. That this information fails to reach those scientists who would benefit from a better understanding of the milieu in which their Soviet counterparts live and work, is perhaps due to its being published primarily in the sociological literature. The merit of the present work is not so much that it provides any fundamentally new material, but that it presents the consensus of present thinking in a readily available form.

This is not to say that the essays have been written with scientists in mind. As Loren R. Graham observes in "Reasons for Studying Soviet Science", the principal reason for compiling this book was the desire to understand better the nature of the Soviet Union as a whole. The emphasis given to the Lysenko affair (which, in Mark Adams's "Science, Ideology and Structure", occupies a major place in the book), or the controversy over the pollution of Lake Baikal of the mid 1960s, is thus not so much because of the scientific importance of the issues, but as examples, on the one hand of how political influence can effectively destroy a whole scientific discipline, and, on the other, of a rare case of public disagreement within Soviet society over a major policy issue. Indeed, at times this emphasis may prove enlightening to the scientist. As Thane Gustafson points out in his essay "Why doesn't Soviet Science do better than it does?", the simple answer which jumps to mind — "political interference" — is not necessarily the whole truth. "Crude ideological or political interference", he says, is the "exception not the rule"; the hierarchical structure and centralized planning system in the Soviet Union, the educational system which is stronger on theoretical than practical training and the uniting of administrative and research decision-

making powers in the same individual must all be taken into account. Whether or not one wholly agrees with his conclusion that the "arrangements developed in the United States are more effective in creating a favorable environment for rapid scientific advance than those of the Soviet Union", his ideas are a useful counter-balance to émigré writers such as Mark Popovsky who attributes the Soviet short-fall in, say, Nobel prizes to political considerations.

How valid it is to judge scientific productivity by such methods is a moot point. Linda L. Lubrano in her essay on scientific collectives does not commit herself; she simply reproduces a number of Soviet tables, assessing productivity in terms of papers written and candidates trained, to illustrate the Soviet concern with norms and output. When she goes on to review Soviet studies on scientific motivation and the psychology of research workers, her exposition lapses into a turgidity and dullness typical of much of the Soviet writing on the subject.

Unfortunately, Professor Lubrano's article was written before the launching of the current press debate in *Literaturnaya Gazeta* on the prestige of pure, as opposed to applied research, many contributions to which have dealt with precisely the same problems of motivation and performance assessment in a far more lively form. This gives a slightly dated air to Lubrano's article, as — although to a lesser extent — it does to the two essays on applied research: Bruce Parrott's "The Organizational Environment of Applied Research" and Kendall E. Bailes's "The Technical Specialists' Social Composition and Attitudes". The latter is almost entirely sociological in content; the former, much wider in scope, ranges over such issues as the changing role of the Academy of Sciences in promoting applied research, the amalgamation of R&D units with production enterprises in the 1970s, and even the degree to which Soviet scientists are alienated from the bureaucratic system.

Taken individually, the seven essays cover a wide range of topics essential to a better understanding of Soviet science. Unfortunately, they remain seven essays, rather than becoming a book in their own right. There is no index, nor are there any useful aids such as a glossary of Soviet acronyms and organizations.

For a scientific audience, the main deficiency of the book is that it was not primarily intended for scientists. Yet, in the absence of a more suitable work, it is to be recommended as background reading to anyone who has dealings not only with individual Soviet scientists, but also with the Soviet scientific establishment. □

Vera Rich is a journalist who specializes in East European affairs.

Volume 3 of *True Visual Magnitude Star Atlas, Northern Stars*, by C. Papadopoulos and C. Scovil (for review see *Nature* 284, 680; 1980) has been published by Pergamon.

Tumour as parasite

P.B. Medawar

The Interaction of Cancer and Host: Its Therapeutic Significance. By Michael F.A. Woodruff. Pp.467. ISBN 0-8089-1265-8. (Grune & Stratton/Academic: 1980.) \$46.50, £26.20.

SIR Michael Woodruff, FRS, is one of a small and important group of surgeons that was brought into being by transplantation therapy: a man equally at home in ward and laboratory who has contributed to both the science and the arts of transplantation. The range of topics under discussion here and the depth of their treatment are such as we normally associate with symposium proceedings or multi-author volumes. Neither is a satisfactory genre for scientific communication, for multi-author works usually have a rootless and untidy character because each author busies himself with the part of the business entrusted to himself, often to the neglect of more general considerations.

The volume under review is open to no such reproaches. It is the work of a single mind — and a very good mind it is too: spaciouly well informed, critical where criticism is called for (witness Woodruff's treatment of the so-called "autonomy" of tumours and the discussion of the notion of the immunological surveillance), but sanguine too in its appraisal of the contribution laboratory science may make to the prophylaxis or treatment of tumours.

A majestic list of references citing titles and first and last page numbers occupies pp.289–419. Over a long period of time such reference lists repay their authors for some of the time and trouble that goes into them but we readers are beneficiaries right away.

In the preparation of a work of this scale and unified structure the time must be set when copy is all packed up and sent to press: no useful purpose is served by delaying publication in the hope that someone will come forward with a revelation without which the book would be incomplete. Woodruff's work is conceptually up to date and its character is such that it matters not the least that it does not include every contribution to tumour biology up to the date of publication of this issue of *Nature*. No reference is made, for example, to the anti-tumour action of retinoids nor to recent epidemiological evidence pointing to the complicity of vitamin A in resistance to tumours. These very recent advances were known only as premonitory rumblings to those who were listening for them. □

Sir Peter Medawar is Head of the Transplantation Biology Section of the Medical Research Council's Clinical Research Centre and is currently engaged in research on tumour prophylaxis.

IQ: the environmentalist position

Richard Lynn

Race, IQ, and Jensen. By James R. Flynn. Pp.313. ISBN 0-7100-0651-9. (Routledge & Kegan Paul: 1980.) £10.50, \$27.50.

HITHERTO, the environmentalist case on the race and IQ issue has been characterized more by emotion and ideology than by reasoning and science. This deficiency is to some degree remedied by the present volume in which Dr Flynn gives a cogent and reasoned exposition of the environmentalist position.

The author recognizes that the chief problem for the environmentalist is to find an environmental factor capable of depressing the mean Negro IQ by one standard deviation or 15 IQ points. He recognizes also that many of the possible candidates frequently advanced by environmentalists do not begin to stand up to examination. This is particularly true of both poverty and discrimination. Among other problems these possibilities collapse in the light of the experience in the United States of the Jews and the Chinese. Both of these ethnic groups arrived as penniless immigrants and suffered initially from the handicaps of poverty and discrimination. Yet both minorities have mean IQs at least as high as those of Caucasians of north-west European descent.

It remains possible, however, that there is a more subtle environmental factor depressing the Negro IQ, or perhaps a set of factors each exerting a relatively small effect. The author proposes low parental IQ, poor nutrition during the mother's pregnancy, lack of stimulation in early childhood, family dislocation, low self-image and poor education. Taken together the author argues that these could be sufficient to account for the phenomenon. The difficulties with these arguments are twofold. First, as the posited deleterious effects are small they are hard to detect. Second, there is a chicken and egg problem in so far as the proposed factors are mainly functions of low IQ in Negro parents.

While there are manifest weaknesses in the environmentalist position, there are also problems in the genetic hypothesis. Two of these are particularly important. In the first place, the genetic hypothesis requires that in individuals of mixed Negro and Caucasian ancestry the IQ should be positively related to the proportion of Caucasian genes. There is contradictory evidence on this point and it is certainly by no means clear that this is the case. Secondly, there is a study of the children of German mothers and American Negro servicemen, in which it was found that the IQ of these children differed little from the mean of the German population.

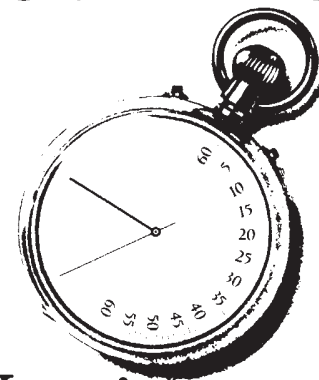
It is a pity that the author has largely confined his discussion to the American data. The results of intelligence testing of racial and ethnic populations in other parts

of the world are an important dimension of the problem and should be considered in any comprehensive discussion. Thus, for instance, if there were a Negro population somewhere in the world with a mean IQ of 100 the environmentalist case would be considerably strengthened. There is a substantial amount of evidence on this question from Negro populations in London, Jamaica and a number of countries in Africa, but in all cases their mean IQ is either about the same as that of American Negroes or else lower. These results cannot be readily explained in terms of test bias, because Mongoloid populations in Japan, Singapore, Taiwan and Hong Kong obtain about the same mean IQ as those of Caucasians in the United States, north-west Europe, Australia and New Zealand. This consistency of intelligence levels across many different nations would appear to suggest a significant genetic determination.

While the present volume is limited in so far as it does not come to terms with this important set of data, it has to be welcomed as a scholarly and informed presentation of the environmentalist position. □

Richard Lynn is Professor of Psychology at the New University of Ulster, Coleraine.

IRCS JOURNALS



Keeping pace with science

IRCS Medical Science
St. Leonard's House
St. Leonardgate,
Lancaster, U.K.

A member of Elsevier Science Publishers

Circle No. 12 on Reader Enquiry Card.

Einstein refined — just in time

W.B. Bonnor

Exact Solutions of Einstein's Field Equations. By D. Kramer *et al.* Pp. 425. ISBN 0-521-23041-1. (Cambridge University Press: 1981.) £30, \$72.

PROGRESS in theoretical physics consists largely of formulating equations and solving them. For instance, the equations put forward by Newton, and by Maxwell, are still being solved for application to special problems; and theorists have been seeking solutions of Einstein's equations since general relativity was discovered in 1915.

Solutions, yes — but why *exact* solutions? Aren't approximate solutions good enough? E. Schrödinger wrote (about Einstein's unified field theory) that "the exact solutions, involving strong fields, have disclosed the ingenuity of the mathematicians who discovered them, but nothing more". Approximate solutions are certainly good enough when a theory is well understood, like Newton's mechanics, and there is no danger of series failing to converge. But in a mathematically complicated theory like general relativity one does not always trust approximations. For example, after half a century of approximate work on motion, there is still no agreement on whether freely falling bodies radiate gravitational waves. A suitable exact solution could settle this once and for all.

A vast number of exact solutions of Einstein's equations have been discovered since 1915. With the half-dozen or so exceptions to be found in nearly all textbooks, these have lain buried in the literature. A warm welcome will therefore be given to this book. The magnitude of the authors' task is indicated by the bibliography which contains about 1,000 references; and these have been sifted from a list of twice as many. Most of the solutions given have been checked, which must have been a monumental task. Indeed, perhaps the work has been done only just in time: had it been left any longer it might not have been possible at all, because papers on exact solutions are appearing at the rate of over 100 per year.

It is possible to use the book simply as a catalogue in which to search for an exact solution, knowing to begin with some of its properties. However it is much more than this. It was essential to impress some order on the plethora of exact solutions obtained by a variety of methods. The authors choose as their principal methods of classification, first, symmetry groups and, second, Petrov types. They give an excellent introduction to both these topics, including much not easily found in other books. Part I, General Methods, also includes chapters on differential and Riemannian geometry, and on the Newman-Penrose formalism.

The authors have done such a fine job that it is important to realize what the book does *not* include. First, as stated in the Introduction, the exact solutions given are those arising from the following energy-momentum tensors: vacuum, electromagnetic fields, pure radiation, dust and perfect fluids. Although all the really important exact solutions fall under these headings, it should be noticed that some areas of physical interest are excluded, for example, charged matter and viscous fluids. It must also be remembered that nearly all solutions in the book are purely local, and no consideration is given to their global existence. Nor are boundary value problems within the authors' terms of reference, so the matching of different

solutions is not considered; this leads to the somewhat bizarre omission of a famous solution, that of Einstein and Strauss for a star embedded in the expanding Universe. Another possible source of misunderstanding for the unwary reader lies in the quite forgivable practice of giving the general form of a solution wherever possible, and omitting reference to special cases, even where these may be of physical importance.

All seriously interested in general relativity should get this book. They should also take advantage of the authors' invitation to inform them of omissions and of new solutions. In this way we may hope that the catalogue will be kept up to date and re-published from time to time. □

W.B. Bonnor is Professor of Mathematics at Queen Elizabeth College, University of London.

Les sciences bien vulgarisées

Alan Isaacs

Concise Encyclopedia of the Sciences. Edited by J.-D. Yule. Pp.590. ISBN 0-87196-491-0. (Facts on File: 1981.) \$29.95.

BEING able to praise one's competitors is an indulgence not frequently available to those of us who earn a living by what the French call the *vulgarisation* of science. However, this is an excellent book, well written, expertly edited and lavishly illustrated mostly in full colour — or rather color, for this is the American edition of a book one version of which has already been published in the UK (Phaidon, 1978; £9.95). At \$29.95 the American edition is undoubtedly good value — for the youngster with an interest in almost any branch of science it will certainly be a most acceptable gift. Uncles (and aunts) please note. It will also be a valuable asset for any library and for readers with no specialized knowledge of science.

Broadly speaking it does exactly what it says it will do — supplies both a dictionary of the most commonly encountered words of science as well as the encyclopaedic background material necessary for understanding their use in a wider context. The 400,000-word text is competent, accurate and jargon-free; for a book of this size it is difficult to see how it could be improved. One third of the text area is devoted to illustrations, almost all of which are a pleasure to look at, both because they communicate the information they are intended to communicate and because they are either well-chosen photographs or well-drawn diagrams. Particularly effective are the historical panels that summarize the history of some subjects and the "context panels" that illustrate the interrelationships between branches of other subjects.

If one has to criticize, one could say that it is perhaps unwise to have claimed (as the jacket does) to have covered philosophy and medicine. In both these fields the coverage is rather less than adequate. In philosophy, for example, the extent of the lacunae can be judged from the fact that there are no entries for ethics or existentialism, nor are Duns Scotus or Aquinas to be found. As Lady Bracknell might have said — to lose one scholastic doctor may be regarded as a misfortune; to lose two looks like carelessness!

In medicine the coverage is more thorough but it is perhaps surprising to find an entry for lymphoma but not one for the more common but less ominous papilloma. Some of the more recent techniques, such as tomography and thermography, are also omitted, as is the now ubiquitous Emiscanner.

If no attempt had been made to cover either philosophy or medicine there would perhaps have been room to deepen the treatment of the growing points of physical science. Quarks, for example, get only a one-sentence mention under subatomic particles, the state of play in nuclear fusion and fast-breeders is not explained, and the alternative energy field is not very well covered (nothing on wave or wind power or biomass energy).

It is, however, uncharitable to dwell on these gaps; so much is included and treated so well. This is undoubtedly an excellent reference book. □

Alan Isaacs is editor of the Macmillan Encyclopedia, the Penguin Dictionary of Science, Longman's New Dictionary of Physics, and science editor of Collins English Dictionary. He has contributed to many other dictionaries and encyclopaedias.